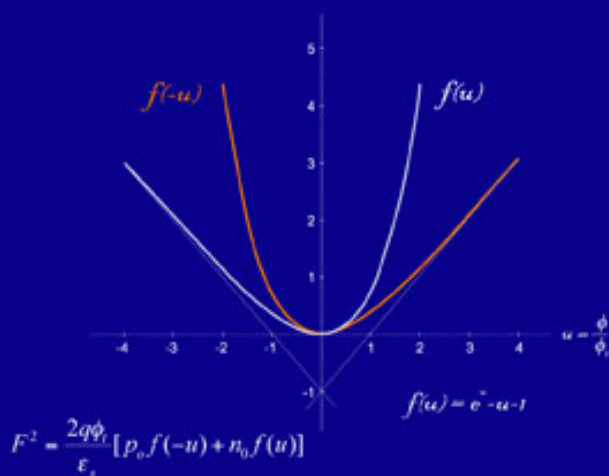


ASSET

International Series on Advances in Solid State Electronics and Technology

Founding Editor: Chih-Tang Sah

MOSFET MODELING FOR CIRCUIT ANALYSIS AND DESIGN



Carlos Galup-Montoro
Márcio Cherem Schneider

World Scientific

MOSFET MODELING FOR CIRCUIT ANALYSIS AND DESIGN

**International Series on Advances in Solid State Electronics and Technology
(ASSET)**

Founding Editor: Chih-Tang Sah

Published:

Modern Semiconductor Quantum Physics

by Li Ming-Fu

Topics in Growth and Device Processing of III-V Semiconductors

by Stephen John Pearton, Cammy R. Abernathy & Fan Ren

Ionizing Radiation Effects in MOS Oxides

by Timothy R. Oldham

MOSFET Modeling for VLSI Simulation: Theory and Practice

by Narain Arora

Forthcoming:

The Physics and Modeling of MOSFETS: Surface-Potential Model HiSIM

by Mitiko Miura-Mattausch, Hans Jürgen Mattausch & Tatsuya Ezaki

BSIM4: Theory and Engineering of MOSFET Modeling for IC Simulation

by Weidong Liu & Chenming Hu

ASSET

International Series on Advances in Solid State Electronics and Technology

Founding Editor: Chih-Tang Sah

MOSFET MODELING FOR CIRCUIT ANALYSIS AND DESIGN

Carlos Galup-Montoro
Márcio Cherem Schneider

Federal University of Santa Catarina, Brazil

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

MOSFET MODELING FOR CIRCUIT ANALYSIS AND DESIGN

International Series on Advances in Solid State Electronics and Technology

Copyright © 2007 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-256-810-6

ISBN-10 981-256-810-7

Disclaimer: This book was prepared by the authors. Neither the Publisher nor its Series Editor thereof, nor any of their employees, assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information. The contents, views, and opinions of the authors expressed herein do not necessarily state or reflect those of the Publisher, its Series Editor, and their employees.

Printed in Singapore.

To Marlene and Rita

This page intentionally left blank

Foreword

The purpose of this compact modeling monograph series is to provide an archival reference on each specific MOS transistor compact model as described by the originators or the veterans of each compact model. The monograph idea came about when this editor was looking into the literature to prepare for a keynote address, invited by the Founder of the Workshop on Compact Modeling, Professor Xing Zhou of Nanyang Technology University, and his program committee, to be presented at its 4th Workshop on May 10, 2005. The topic was on the history of MOS transistor compact modeling, a subject this editor could not find a reference or book that provided the descriptions of each of the dozen or more MOS transistor compact models, which had been extensively developed for the first-generation computer-aided circuit design applications during 1995-2005, such as the use of the Berkeley BSIM and SPICE. A second purpose is to serve as textbooks for graduate students and reference books for practicing engineers, to rapidly distribute the detailed design methodologies and underlying physics in order to meet the ever faster advances in the design of silicon semiconductor MOS and bipolar-junction-transistor integrated circuits, which contain hundreds or thousands of transistors per circuit or circuit function. I am especially thankful to the authors of the four startup monograph volumes who concurred with me and agreed to take up the chore to write their books in the very short time of less than six months in order to be published in one year, which we try as a rapid response to document the latest advances. It is also the objective of this monograph series to provide timely updates via website exchanges between the readers and authors, for public distribution, and for new editions when sufficient materials are accumulated by the authors.

We are very pleased to publish the graduate course lecture notes taught by Professors Carlos Galup-Montoro and Márcio C. Schneider of the Federal University of Santa Catarina, Brazil, as the first monograph of original unpublished work of this series on MOS transistor compact modeling. It is the first textbook and reference book using the next generation surface potential approach to develop compact MOS transistor models by ways of the inversion charge in the surface channel of the conduction carriers, electron (hole) channel for the n-channel (p-channel) MOS transistor. It starts out with a concise definition of the inversion charge in terms of the integrals of the surface potential variable which no other publications have provided as the starting point. This provides the mathematical rigor on the meaning of the inversion charge.

I would like to thank all the WSPC editors and this monograph volume's copyeditor Mr. Steven Patt (Singapore), led by Dr. Yubing Zhai (New Jersey) for their timely efforts, and Professor Kok-Khoo Phua, Founder and Chairman of WSPC, for his support, all of which have made it possible to attain a less-than-one-year turn-around time to print each monograph volume, in order to meet our intention of responding to the rapid advances of the state of the art of computer-aided integrated circuit design.

Chih-Tang Sah

Gainesville, San Diego, Singapore

December 31, 2006

This page intentionally left blank

Preface

Complementary metal-oxide semiconductor (CMOS) technology is the leading electronics technology and will continue to be for the next few years. The metal-oxide-semiconductor field-effect transistor (MOSFET) is the basic building block in CMOS technologies and, as a consequence, the predominant device of integrated circuits. Therefore, MOSFET modeling plays an important role in interfacing the circuit design community with the device and technology community.

In order to be of practical usefulness for circuit designers, MOSFET models should be compact, *i.e.*, they should provide efficient and accurate algorithms to calculate charges, currents, and their derivatives. The compact modeling of MOS transistors for integrated circuit design has, for many years, been driven by the needs of digital circuit simulation. However, the trend which started in the late 1970's toward mixed analog-digital design generated the necessity for MOSFET models appropriate for analog and radio-frequency (RF) design as well. As a result, two types of advanced models are currently considered as adequate for circuit analysis and design, inversion charge and surface potential-based models. In the former approach, currents and charges are expressed in terms of the inversion charge densities at the source and drain ends of the transistor channel. In the latter, drain current and charges are related to the surface-potential at the ends of the channel.

This book provides both an overview of the basic physics theory required to build compact MOSFET models and a unified treatment of inversion-charge and surface-potential models. The text presents a fresh view of compact modeling, having completely abandoned the regional modeling approach. Regional models are presented as asymptotic cases of the all-region model, with the main objective of developing the understanding of the reader regarding transistor operation in that particular region. Compact expressions for hand analysis or for automatic

synthesis, valid in all the operating regions, are presented throughout the book. Most of the accurate expressions for computer simulation used in the new generation compact models are derived.

The book starts with a short introduction to MOSFET transistor modeling. Chapter 2 presents the two- and three-terminal MOS structures. Because it forms the basis of MOS compact models, the equivalent capacitive circuit of the MOS structure is carefully derived. A rigorous definition of pinch-off based on charge is given. Chapter 2 includes a thorough treatment of the unified charge control model (UCCM) which is at the center of the so-called charge-based models. The long-channel MOSFET theory is the subject of Chap. 3. Compact drain current models based on either surface potential or inversion charge are rigorously derived from the exact Pao and Sah double-integral formula. The chapter finishes with the introduction of a current-based model appropriate for design and parameter extraction.

In Chap. 4, the basic long-channel theory is extended to obtain practical compact models. Of particular relevance is the definition of saturation in terms of charge. Several small-dimension effects are presented and compact models are developed for them. Chapter 5 deals with stored charges and capacitive coefficients, taking into account short-channel effects. The complete small-signal model for quasi-static operation is presented.

The book strongly emphasizes the modeling of fluctuations, introducing a unified approach for both space (matching) and time (noise) fluctuations. Chapter 6 is dedicated to mismatch modeling and Chap. 7 aims at giving the reader an in-depth understanding of thermal and low-frequency noise modeling.

Non-quasi-static models for high-frequency applications are developed in Chap. 8. Compact models for gate and bulk currents originating from tunneling and high-field effects are developed in Chap. 9. Advanced MOSFET structures with their specific models are presented in Chap. 10. A simple but accurate procedure for parameter extraction is introduced in Chap. 11. The approaches taken by the developers of the main next generation compact models are summarized in Chap. 12. Finally, some fundamental concepts essential for the

development of the theory behind MOS transistors are summarized in the appendices.

This textbook has been organized to fulfill the needs of several types of audience. It is intended as a key reference on MOSFET modeling for graduate courses as well as senior level courses. For short courses, different selections of topics are recommended. A course emphasizing general principles should be based on Chaps. 1 to 5. A course for digital designers may cover Chaps. 1 to 6 and Chap. 9, plus selected sections from Chap. 10. Since RF circuits are the most demanding regarding models, a course for RF designers should cover the complete book, with the possible exception of Chap. 10 on advanced MOS structures.

The authors would like to acknowledge those individuals who contributed to the writing of this book and to the research on which it is based. We are very grateful to Professor Chih-Tang Sah, who kindly requested that we write this book. A book chapter about the all-region one-equation MOS transistor that we wrote several years ago was the first step on this path. We are indebted to Professor Edgar Sánchez-Sinencio, who invited us to write that chapter.

We are also very grateful to our former and current students who developed the research on MOS modeling: Dr. Ana Isabela Araújo Cunha, who made the first and fundamental contributions to the charge-based modeling approach in our group; Dr. Oscar da Costa Gouveia Filho, who developed the computer-implemented model, the Advanced Compact MOSFET (ACM) model; Dr. Alfredo Arnaud, who started our work on the unified modeling of fluctuations (noise and mismatch) and its application to low-power circuit design; Hamilton Klimach, who has been developing the modeling and characterization of mismatch in MOSFETs; Viriato Correa Pahim, who contributed to the comparison between the inversion charge and surface potential approaches and also to noise modeling; Márcio Bender Machado, for improving the parameter extraction methodology.

We thank Dr. Siobhan Wiese, who reviewed the text to reduce the distance from our English to the language of her homeland. We are also very grateful to Paula Chakkour for preparing the figures.

Finally, the support of CNPq and CAPES, Brazilian agencies for scientific development, was invaluable to the research in our laboratory that resulted in this book.

Florianópolis, April 2006

Carlos Galup-Montoro and Márcio Cherem Schneider

Contents

Foreword	vii
Preface	ix
List of Selected Symbols	xix
 Chapter 1 Introduction	 1
1.1 MOS integrated circuits	1
1.2 Compact MOSFET models for circuit analysis and design	2
1.3 A brief history of compact MOS transistor models	3
Bibliography	5
 Chapter 2 The MOS Capacitor	 7
2.1 Equilibrium electron and hole concentrations	7
2.2 The field effect in bulk semiconductors	9
2.2.1 Fundamental considerations	9
2.2.2 Poisson-Boltzmann equation in bulk semiconductors	10
2.3 The ideal two-terminal MOS structure	14
2.3.1 Small signal equivalent capacitive circuit	17
2.3.2 Ideal C-V plots	21
2.3.3 Effect of work function difference	26
2.3.4 The flat-band voltage	29
2.4 The three-terminal MOS structure with uniformly doped substrate	32
2.4.1 Quasi-equilibrium electron and hole concentrations	32
2.4.2 Surface potential master equation	35
2.4.3 Small-signal equivalent capacitive circuit in inversion	38
2.4.4 Basic approximations for compact modeling	41
2.4.4.1 The charge-sheet approximation	41
2.4.4.2 The approximate linear relationship between inversion charge density and surface potential	43

2.4.5	The pinch-off voltage	45
2.4.6	The unified charge control model (UCCM)	47
2.4.7	Comparison between charge-sheet model and UCCM	49
2.5.	Real C-V curves: Interface traps, polysilicon depletion, and quantum effects	51
2.5.1	Interface-trap capacitance	52
2.5.2	Polysilicon depletion	54
2.5.3	Quantum mechanical (QM) effects on the semiconductor capacitance	57
	Bibliography	61
	Problems	64

Chapter 3	The Long-Channel MOSFET: Theory and dc Equations	67
3.1	A brief history of MOSFET theory	67
3.2	MOSFET operation	71
3.3	The Pao-Sah exact I-V model	73
3.4	Compact surface potential MOSFET models	79
3.4.1	Linearized surface potential models	83
3.4.2	Brews-van de Wiele charge-sheet formula for the current	84
3.5	Charge control compact model	85
3.6	Comparison between models	86
3.7	Design-oriented MOSFET model	88
3.7.1	Forward and reverse components of the drain current	88
3.7.2	Asymptotic behavior of the drain current in weak and strong inversion	91
3.7.3	Universal dc characteristics	94
3.7.4	Small-signal transconductances	98
3.7.5	The transconductance-to-current ratio	101
3.7.6	Small-signal MOSFET model at low frequencies	103
	Bibliography	105
	Problems	107

Chapter 4	The Real MOS Transistor: dc Models	113
4.1	Effective mobility	113
4.2	Velocity saturation	119
4.3	Saturation charge and saturation voltage	122
4.4	Channel length modulation	127

4.5	Effect of source and drain resistances	129
4.6	Ballistic transport	129
4.7	Short and narrow channel effects	131
4.7.1	Short-channel effect	131
4.7.2	Reverse short-channel effect	134
4.7.3	Narrow-channel effect	135
4.7.4	Drain-induced barrier lowering	137
4.7.5	Temperature effect on the threshold voltage	139
4.8	Impact of small geometry effects on transistor model	140
4.9	Benchmark tests for dc MOSFET models for circuit simulation	144
4.9.1	Series-parallel association of transistors	145
4.9.2	Gummel symmetry test	149
	Bibliography	151
	Problems	154
Chapter 5	Stored Charges and Capacitive Coefficients	157
5.1	Transient analysis of the MOS transistor	158
5.1.1	The continuity equation	158
5.1.2	Definitions of source and drain charges	159
5.2	Quasi-static charge-conserving model	163
5.2.1	Stored charges	163
	5.2.1.1 Symbolic MOSFET model	164
	5.2.1.2 Stored charge calculations	165
5.2.2	Transit time	170
5.2.3	Capacitive coefficients	171
	5.2.3.1 Intrinsic small-signal models	173
	5.2.3.2 Calculation of the capacitive coefficients	175
	5.2.3.3 Charge conservation and transcapacitances	183
	5.2.3.4 Intrinsic transition frequency	185
5.3	Charges and capacitances of the extrinsic transistor	188
5.4	Small-dimension effects on charges and capacitances	190
	5.4.1 Effect of velocity saturation on stored charges	190
	5.4.2 Effect of velocity saturation on capacitances	194
	5.4.3 Other small-dimension effects on capacitances	199
	Bibliography	200
	Problems	201

Chapter 6	Mismatch Modeling	203
6.1	Introduction	203
6.2	Random mismatch in MOS transistors	204
6.3	Consistent model for drain current fluctuation	206
6.4	Number fluctuation mismatch model	207
6.5	Specific current mismatch	212
6.6	Mismatch model in terms of inversion level	213
6.7	Experimental results and discussion	215
	Bibliography	222
	Problems	224
Chapter 7	Noise in MOSFETs	227
7.1	Sources of noise	227
7.1.1	Thermal noise	227
7.1.2	Shot noise	229
7.1.3	Flicker noise	231
7.2	Noise modeling using the impedance field method	232
7.2.1	Thermal noise in MOSFETs	236
7.2.2	Flicker noise in MOSFETs	236
	7.2.2.1 Measurement results	240
7.3	Consistency of noise models	243
7.4	Design-oriented noise models	246
7.4.1	Relationship between thermal noise and transconductance	246
7.4.2	Flicker noise in terms of inversion levels	247
7.4.3	The corner frequency	250
7.5	Small-dimension effects on MOSFET noise	251
7.5.1	Review of the effect of velocity saturation on the drain current	251
7.5.2	The impedance field for short channel MOSFETs	253
7.5.3	Drain current noise of short channel MOSFETs	255
	Bibliography	260
	Problems	262
Chapter 8	High-Frequency Models	265
8.1	Introduction	265
8.2	Non quasi-static operation of the MOSFET	266
8.3	Non-quasi-static small-signal model	267
8.4	The y-parameter model	275

8.5	Channel segmentation	280
8.6	Induced gate noise	281
	Bibliography	290
	Problems	292
Chapter 9	Gate and Bulk Currents	293
9.1	Gate tunneling current	293
9.1.1	Tunneling through a potential barrier	294
9.1.2	Compact model for tunneling in MOS structures	295
9.1.2.1	Impact frequency	297
9.1.2.2	Equations for the tunneling current	299
9.1.2.3	Comparison between models	301
9.1.2.4	Partition of the gate current into source and drain components	303
9.2	Bulk current	306
9.2.1	Gate induced drain leakage	306
9.2.2	Impact ionization current	308
9.2.2.1	Weak avalanche	308
9.2.2.2	Breakdown	313
	Bibliography	315
	Problems	317
Chapter 10	Advanced MOSFET Structures	319
10.1	Introduction	319
10.2	Deep submicron planar MOS transistor structures	320
10.3	Silicon-on-insulator (SOI) CMOS transistors	324
10.3.1	Pinch-off voltage calculation	334
10.4	Undoped surrounding-gate transistors	338
10.5	Multiple-gate transistors	343
	Bibliography	345
	Problems	348
Chapter 11	MOSFET Parameter Extraction	349
11.1	Introduction	349
11.2	Threshold voltage and short-channel effects	350
11.3	Specific current and effective channel length and width	355
11.4	Slope factor and subthreshold slope	357
11.5	Mobility degradation with transversal field	359

Bibliography	363
Problems	364
Chapter 12 Advanced MOSFET Models for Circuit Simulators	367
12.1 Surface potential- vs. inversion charge-based models	368
12.2 Charge-based models	369
12.2.1 The ACM model	370
12.2.2 The EKV model	373
12.2.3 The BSIM5 model	375
12.3 Surface potential models	376
12.3.1 The HiSIM model	378
12.3.2 MOS model 11	380
12.3.3 The SP model	382
Bibliography	383
Problems	386
Appendix A Electrostatics in One Dimension	387
Appendix B Electrostatics in Semiconductors	391
Appendix C Drift-diffusion Current Model	393
Appendix D Continuity Equations	397
Appendix E Basics of pn Junctions	399
E.1 Electrostatics in equilibrium	400
E.2 Electrostatics of forward- and reverse-biased junction	401
E.3 Quasi-Fermi potentials in the pn-junction	402
Appendix F Hall-Shockley-Read (HSR) Statistics	405
Appendix G Interface Trap Capacitance	409
Index	415

List of Selected Symbols

Symbol	Description	Unit
A_{ISQ}	Mismatch factor of the sheet normalization current	%- μm
C'_b	Depletion capacitance per unit area	F/cm ²
C_{bd}	Bulk-drain capacitance	F
C_{bs}	Bulk-source capacitance	F
C'_c	Semiconductor capacitance per unit area	F/cm ²
C_{ds}	Drain-source transcapacitance	F
C'_{fb}	Flat-band capacitance per unit area	F/cm ²
C_{gb}	Gate-bulk capacitance	F
C'_{gb}	Gate-bulk capacitance per unit area	F/cm ²
C_{gd}	Gate-drain capacitance	F
C_{gs}	Gate-source capacitance	F
C'_i	Inversion capacitance per unit area	F/cm ²
C_m	$C_{dg} - C_{gd}$	F
C_{mx}	$C_{bg} - C_{gb}$	F
C_{ox}	Total oxide capacitance ($WL_e C'_{ox}$)	F
C'_{ox}	Oxide capacitance per unit area	F/cm ²
C_{sd}	Source-drain transcapacitance	F
C'_{ss}	Interface trap capacitance per unit area ($C'_{ss} = qN'_{ss}$)	F/cm ²
D_n	Electron diffusion coefficient	cm ² /s
D_p	Hole diffusion coefficient	cm ² /s
E_C	Energy at the conduction-band edge	eV
E_F	Fermi energy	eV
E_g	Silicon band gap energy (= 1.12 eV @ 300K)	eV
E_i	Intrinsic Fermi level	eV
E_V	Energy at the valence-band edge	eV
F	Electric field	V/cm
F_{ave}	Average electric field in the transversal (x) direction	V/cm
F_C	Critical electric field	V/cm

f_c	Corner frequency	Hz
F_s	Surface electric field	V/cm
f_T	Intrinsic cutoff frequency	Hz
g_{mb}	Bulk transconductance	A/V
g_{md}	Drain transconductance	A/V
g_{mg}	Gate transconductance	A/V
g_{ms}	Source transconductance	A/V
G_n	Generation rate for electrons	$\text{cm}^{-3}\text{-s}^{-1}$
G_p	Generation rate for holes	$\text{cm}^{-3}\text{-s}^{-1}$
g_0	Output conductance in saturation	A/V
h	Planck's constant ($= 6.63 \times 10^{-34}$ J-s)	J-s
I_B	Large-signal bulk current	A
i_b	Small-signal bulk current	A
I_D	Large-signal drain current	A
i_D	Normalized drain current (I_D/I_S)	None
i_d	Small-signal drain current	A
I_{Dsat}	Saturation drain current	A
i_{Dsat}	Normalized saturation drain current (I_{Dsat}/I_S)	None
I_F	Forward saturation current	A
i_f	Forward normalized current or inversion level at source (I_F/I_S)	None
I_G	Large-signal gate current	A
i_g	Small-signal gate current	A
I_{GD}	Gate-to-drain current	A
I_{GS}	Gate-to-source current	A
I_R	Reverse saturation current	A
i_r	Reverse normalized current or inversion level at drain (I_R/I_S)	None
I_S	Normalization current ($(W/L)\mu_n C'_{ox} n\phi_t^2 / 2$)	A
i_s	Small-signal source current	A
I_{SQ}	Sheet normalization current ($\mu_n C'_{ox} n\phi_t^2 / 2$)	A
$i_{\Delta A}$	Local current fluctuation	A
J_n	Electron current density	A/cm^2
J_p	Hole current density	A/cm^2
k	Boltzmann's constant ($= 1.38 \times 10^{-23}$ J/K)	J/K
L	Channel length	cm
L_D	Extrinsic Debye length	cm
L_e	Effective channel length ($L_e = L - \Delta L$)	cm
n	Free electron concentration	cm^{-3}
n	Slope factor ($1 + C'_b / C'_{ox}$)	None

N^*	Channel carrier density at pinch-off ($-Q'_{IP} / q$)	cm^{-2}
N_A	Acceptor concentration	cm^{-3}
N_D	Donor concentration	cm^{-3}
n_i	Si intrinsic carrier concentration ($= 1.2 \times 10^{10} \text{ cm}^{-3}$ @ 300K)	cm^{-3}
N_{oi}	Equivalent impurity density for mismatch calculation	cm^{-2}
N_{ot}	Equivalent density of oxide traps for 1/f noise calculation	cm^{-2}
N'_{ss}	Density of interface traps per unit potential per unit area	$\text{V}^{-1}\text{-cm}^{-2}$
$N_t(E)$	Density of oxide traps per unit energy per unit volume	$\text{eV}^{-1}\text{-cm}^{-3}$
p	Hole concentration	cm^{-3}
q	Electronic charge ($= 1.60 \times 10^{-19} \text{ C}$)	C
Q_B	Total depletion charge	C
Q'_B	Depletion charge per unit area	C/cm^2
Q'_{Ba}	Depletion charge density deep in weak inversion	C/cm^2
Q_C	Total semiconductor charge	C
Q'_C	Semiconductor charge per unit area	C/cm^2
Q_D	Drain stored charge	C
Q'_F	Virtual charge density at source excluding v_{sat} ($Q'_F = Q'_{IS} - nC'_{ox}\phi_t$)	C/cm^2
Q_G	Total gate charge	C
Q'_G	Gate charge per unit area	C/cm^2
Q_I	Total inversion charge	C
Q'_I	Inversion charge per unit area	C/cm^2
q'_{ID}	Normalized inversion charge density at drain (Q'_{ID} / Q'_{IP})	None
Q'_{IDsat}	Inversion charge density at the saturated drain end	C/cm^2
q'_{IDsat}	Normalized inversion charge density at the saturated drain end (Q'_{IDsat} / Q'_{IP})	None
Q'_{IP}	Pinch-off charge per unit area ($-nC'_{ox}\phi_t$)	C/cm^2
q'_{IS}	Normalized inversion charge density at source (Q'_{IS} / Q'_{IP})	None
Q'_{It}	Virtual charge density excluding v_{sat} ($Q'_{It} = Q'_I - nC'_{ox}\phi_t$)	C/cm^2

Q'_R	Virtual charge density at drain excluding v_{sat} ($Q'_R = Q'_{ID} - nC'_{ox}\phi_t$)	C/cm ²
Q_S	Source stored charge	C
Q'_V	Virtual charge density ($Q'_I + Q'_{IP} - Q'_{IDsat}$)	C/cm ²
Q'_{VD}	Virtual charge density at drain	C/cm ²
Q'_{VS}	Virtual charge density at source	C/cm ²
R_n	Recombination rate for electrons	cm ⁻³ -s ⁻¹
R_p	Recombination rate for holes	cm ⁻³ -s ⁻¹
S_{Id}	Power spectral density of the drain current	A ² /Hz
S_{th}	Thermal noise power spectral density	A ² /Hz
T	Absolute temperature	K
t_{ox}	Oxide thickness	cm
u	Normalized electrostatic potential (ϕ/ϕ_t)	None
u_F	Normalized Fermi potential (ϕ_F/ϕ_t)	None
u_n	Normalized quasi-Fermi potential for electrons (ϕ_n/ϕ_t)	None
u_p	Normalized quasi-Fermi potential for holes (ϕ_p/ϕ_t)	None
V_A	Early voltage	V
V_{ACLM}	Channel length modulation (CLM) component of the Early voltage	V
V_{ADIBL}	Drain induced barrier lowering (DIBL) component of the Early voltage	V
V_C	Channel voltage	V
V_D	Drain voltage	V
v_d	Drift velocity	cm/s
V_{DSsat}	Drain-source saturation voltage	V
V_{FB}	Flat-band voltage	V
V_G	Gate voltage	V
V_P	Pinch-off voltage	V
V_S	Source voltage	V
v_{sat}	Saturation velocity	cm/s
v_{th}	Thermal velocity	cm/s
V_{TO}	Equilibrium threshold voltage	V
W	Channel width	cm
x	Distance in direction perpendicular to the surface	cm
x_i	Inversion channel depth	cm
X_J	Junction depth	cm
y	Distance along the channel	cm

α	Saturation coefficient (Q'_{VD} / Q'_{VS})	None
α_θ	Scattering constant	cm/V
γ	Body effect factor	$V^{1/2}$
Δf	Bandwidth	Hz
ΔL	Channel length shortening	cm
δI_D	Time fluctuation of the drain current	A
ε_{ox}	Permittivity of silicon dioxide (= 3.45×10^{-13} F/cm)	F/cm
ε_s	Permittivity of silicon (= 1.04×10^{-12} F/cm)	F/cm
ζ	Short channel parameter ($\phi_t / (L.F_C)$)	None
η	$(q'_{ID} + 1) / (q'_{IS} + 1)$	None
θ	Mobility degradation factor	$V^{-1/2}$
μ_{eff}	Effective mobility	$\text{cm}^2/\text{V-s}$
μ_n	Electron mobility	$\text{cm}^2/\text{V-s}$
μ_p	Hole mobility	$\text{cm}^2/\text{V-s}$
μ_0	Low field mobility	$\text{cm}^2/\text{V-s}$
ρ	Charge density	C/cm^3
σ_{ID}^2	Squared standard deviation of the drain current	A^2
τ	Channel transit time	s
τ_1	Main time constant of the non quasi-static model	s
ϕ	Electrostatic potential	V
ϕ_{bi}	Built-in potential of a pn junction	V
ϕ_F	Fermi potential	V
ϕ_n	Quasi-Fermi potential for electrons	V
ϕ_p	Quasi-Fermi potential for holes	V
ϕ_s	Surface potential	V
ϕ_{s0}	Surface potential at source	V
ϕ_{sa}	Surface potential deep in weak inversion	V

ϕ_{sL}	Surface potential at drain	V
ϕ_{sp}	Surface potential at pinch-off	V
ϕ_t	Thermal voltage ($kT/q = 26$ mV @ 300K)	V
ω	Angular frequency	rad/s
ω_T	Intrinsic transition angular frequency	rad/s

Chapter 1

Introduction

In this chapter, compact models for the metal-oxide-semiconductor (MOS) transistor are briefly introduced and their evolution since the late 1960s is summarized.

1.1 MOS integrated circuits

During the initial years, from the early 1960s until the late 1970s, the design of integrated circuits (IC) was the domain of circuit designers working within semiconductor companies. This paradigm started to change in the early 1980s, after the introduction of the text “Introduction to VLSI Systems”, by Carver Mead and Lynn Conway [1]. The Mead-Conway text was written to fill a gap in the literature and to introduce electrical engineering and computer science students to integrated system architecture and design. Instrumental in the wide diffusion of the IC design capabilities were the multiproject chip initiatives which began in the 1980s. The access to fast prototyping of chips has allowed engineers other than those working within semiconductor companies to design VLSI chips.

During the 1970s, nMOS was the dominant technology for highly complex digital circuits. In the early 1980s, CMOS became the technology of choice for general-purpose integrated circuit applications [2]. Advantages of CMOS technology include low static power consumption, simple laws of scalability, and stability of operation [2]. Finally, the CMOS technology has proven to be useful not only for digital circuits but also for analog and RF circuits; as a result, many chips

today are complete systems thanks to the flexibility and scalability of the CMOS technology.

As a consequence of the explosive growth of the IC business, including fab and fabless companies, a large number of circuit designers work without having regular interaction with the process and manufacturing engineers. The designers rely on the simulation of their design before building a prototype. To simulate a circuit, simulators make use of element models, which provide a mathematical description of the element behavior in the circuit. The compact MOSFET models provide most of the designers with the essential information concerning electrical properties of the components associated with the manufacturing process of the chip.

1.2 Compact MOSFET models for circuit analysis and design

Computer aided design (CAD) tools are essential elements for circuit design. The productivity of circuit designers is intimately associated with the efficiency of the available arsenal of CAD tools. One of the most useful IC design tools is the circuit simulator, which allows users to enhance the understanding of a circuit and some fine details of its operation. The judicious use of electrical simulators allows the quick evaluation of the circuit performance without the burden of a costly integrated prototype. One must bear in mind, however, that the accuracy of the results provided by the simulator depends on the quality of the circuit element models. Thus, MOSFET models, which serve as the critical communication vehicle between circuit designers and silicon foundries [3], play a crucial role in chip design productivity.

At this point, we differentiate between two categories of device models [4], [5], namely numerical device simulation models and compact models. Numerical device simulators are used to study the device physics and to predict the electrical, optical, and thermal behavior of a device. Numerical device simulators solve a set of (partial differential) equations associated with the physics involved in device operation; their requirements of intensive computation and huge amounts of memory prevent them from being used for circuit simulation. On the

other hand, compact models or equivalent circuit models describe the terminal properties of the device by means of a simplified set of equations or by an equivalent circuit model. Of course, device simulators can be very helpful in the task of determining the equivalent circuit of a device.

This book deals with compact models of the MOS transistor for circuit analysis and design. The purpose of a compact model is to obtain simple, fast, and accurate representations of the device behavior. Compact transistor models are needed to evaluate the performance of integrated circuits containing a large quantity of transistors, sometimes several thousands. MOSFET compact models fall into three categories: (1) Physical models, based on device physics, (2) Table lookup models, in the form of tables containing device data for different bias points, and (3) Empirical models, which represent the device characteristics through equations that fit data. Physical models take considerable time to develop but once they become mature, their advantages are significant: parameters have physical meaning, effects of device geometry, technological parameters, and temperature can be accounted for, statistical modeling can be applied to predict ranges of expected performance [6] and, in many cases, the model can be applied to different generation technologies by simply changing parameters. Throughout this book, we will mainly deal with physical models. We strongly believe that circuit designers who have a background in the physical phenomena behind transistor operation are more capable of quickly analyzing a circuit topology, understanding its limitations, and more easily interpreting simulation results, besides being more ingenious and creative in their designs.

1.3 A brief history of compact MOS transistor models

In the late 1960s and early 1970s, simulation programs aimed at analyzing nonlinear circuits began to be developed. The purpose of this development was mainly to test new circuits made available by the nascent field of integrated circuits. SPICE (Simulation Program with Integrated Circuit Emphasis), released in 1972, is the most tangible result

of the effort in developing simulation programs at UC Berkeley [7]. Since its beginning, SPICE or one of its many derivatives has been an invaluable resource in evaluating IC performance prior to its integration [8].

Essential to SPICE-like simulators are the element models, and, in particular, a compact model for the MOS transistor.

The first MOSFET model for the SPICE circuit simulator, the Level 1 model, often called Shichman-Hodges model [9], is a simplified first-order model suitable for long-channel transistors only. Level 1 describes the current dependence on voltages for gate voltages greater than the threshold voltage; the sub-threshold current is assumed to be zero. In addition, the terminal capacitances, which are described by the Meyer model [10], are not charge-conserving.

The Level 2 model addresses second-order effects associated with small-geometry devices. Unlike in Level 1, the sub-threshold current is not equal to zero. The capacitive model can be either the Meyer model [10] or the Ward-Dutton model [11], the latter conserving charges. Level 2 is computationally very complex and convergence problems are often encountered [5], [12]. The many drawbacks of Level 2 are extensively commented on in [3].

The Level 3 model is a semi-empirical model developed to address the shortcomings of Level 2. It runs faster than Level 2 and convergence problems are seldom encountered. The capacitive model of Level 3 is the Ward-Dutton model. Failure to properly model the sub-threshold current and the output conductance are two of the major drawbacks of Level 3.

The rapid evolution of the MOS technology in the 1980s showed that the Level 1, 2, and 3 models were clearly not appropriate to simulate efficiently circuits with a large number of ever-smaller transistors. A different modeling philosophy to that employed for first generation models was then adopted. BSIM (Berkeley Short-Channel IGFET Model) [13] inaugurated a second generation of MOS transistor modeling, which put less effort into developing physical models but instead concentrated on mathematics for faster and more robust circuit simulation [3]. Convergence problems and negative output conductance were some of the problems that inhibited the use of BSIM for analog designs. Two later developments, BSIM2 and HSPICE Level 28, with

comprehensive modifications of BSIM, made second-generation simulators suitable for analog IC design [12]. The most important shortcomings of second-generation models are their empirical and complex implementation, which incorporates several parameters without clear physical meanings [12].

BSIM3 and its extension BSIM4, along with MOS Model 9, brought into the public domain by Philips, began the third-generation approach [12] of the 1990s, which reintroduced the physical basis into the models. The use of smoothing functions in third-generation models provides continuous and smooth behavior of device characteristics across all the operating regions. Even though it follows the general trend of third-generation models, the EKV model [14] is a fully symmetrical model which uses the bulk voltage as the reference.

The first, second, and third-generation models previously mentioned were the most used for integrated circuit design over the 30-year period, from 1970 to 2000. In these models, the current and charges are approximated by explicit functions of the terminal voltages. The approximate solutions are obtained by interpolating models that are only valid in particular regions of operation. This leads to inaccuracy between operating regions and, thus, inaccuracy in the simulation of circuits where transistors operate between these regions [15]. A considerable effort in academia and companies has been made in the last few years to provide the design community with better compact models that can lead to more accurate simulation results. The development of more accurate models, together with the physics behind them, and a review of some landmark papers concerning compact MOSFET models will be the main subjects of this textbook.

Bibliography

- [1] Carver Mead and Lynn Conway, *Introduction to VLSI Systems*, Addison–Wesley, Reading, Mass., 1980.
- [2] Chih-Tang Sah, “Evolution of the MOS transistor – From conception to VLSI,” *Proc. IEEE*, vol. 76, no. 10, pp. 1280–1326, Oct. 1988.
- [3] Daniel P. Foty, *MOSFET Modeling with SPICE - Principles and Practice*, Prentice Hall, Upper Saddle River, NJ, 1997.

- [4] Henk C. de Graaff and François M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Wien, 1990.
- [5] Narain Arora, *MOSFET Models for VLSI Circuit Simulation Theory and Practice*, Springer-Verlag, Wien, 1993.
- [6] Yannis Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. , McGraw-Hill, Boston, 1999.
- [7] Andrei Vladimirescu, *The SPICE Book*, John Wiley & Sons, New York, 1994.
- [8] Kenneth S. Kundert, *The Designer's Guide to SPICE & SPECTRE*, Kluwer Academic Publishers, Boston, 1995.
- [9] Harold Shichman and David A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," *IEEE J. Solid-State Circuits*, vol. 3, no. 5, pp. 285-289, Sep. 1968.
- [10] John E. Meyer, "MOS models and circuit simulation," *RCA Review*, vol. 32, pp. 42-63, Mar. 1971.
- [11] Donald E. Ward and Robert W. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE J. Solid-State Circuits*, vol. 13, no. 5, pp. 703-708, Oct. 1978.
- [12] William Liu, *MOSFET Models for SPICE Simulation, Including BSIM3v3 and BSIM4*, John Wiley & Sons, New York, 2001.
- [13] Bing J. Sheu, Donald L. Scharfetter, Ping-Keung Ko, and Min-Chie Jen, "BSIM: Berkeley Short-Channel IGFET Model for MOS transistors," *IEEE J. Solid-State Circuits*, vol. 22, no. 4, pp. 558-566, Aug. 1987.
- [14] Christian C. Enz, François Krummenacher, and Eric A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *J. Analog Integr. Circuits Signal Process.*, vol. 8, pp. 83-114, July 1995.
- [15] Josef Watts, Colin McAndrew, Christian Enz, Carlos Galup-Montoro, Gennady Gildenblat, Chenming Hu, Ronald van Langevelde, Mitiko Miura-Mattausch, Rafael Rios, Chih-Tang Sah, "Advanced compact models for MOSFETs," *Proc. Workshop on Compact Modeling, Nanotech 2005*, pp. 3-12.

Chapter 2

The MOS Capacitor

The metal-oxide-semiconductor (MOS) structure is the core of the MOS technology. The two-terminal structure is called the MOS capacitor and as a stand-alone device it plays a fundamental role in technology characterization and parameter extraction. Also, it is sometimes used as a circuit element to implement a varactor. Important as they are, these uses are not the subject of this chapter. Here, we will focus on the operation of the MOS capacitor as the core of the transistor (four-terminal device). The two- and three-terminal MOS structures are presented.

Because it forms the basis of MOS compact models, the equivalent capacitive circuit of the MOS structure is carefully derived. A rigorous definition of pinch-off based on charge is given. This chapter includes a thorough treatment of the unified charge control model (UCCM), which is at the center of the so-called charge-based models.

2.1 Equilibrium electron and hole concentrations

The first step in the development of a model for the current or the charges in a semiconductor device is the calculation of the carrier concentrations in the device. The interest in the equilibrium condition derives from the fact that it not only serves as a reference state, but often semiconductor devices can be considered to be operating in a quasi-equilibrium regime.

In equilibrium (also referred to as thermal or thermodynamic equilibrium) electrons and holes in an ideal crystalline semiconductor behave as ideal gases if the doping concentrations are not too high (typically of the order of 10^{19} cm^{-3} or less), as is the case for the bulk

material of MOS devices. Consequently, electrons and holes follow Boltzmann's law [1] and their concentrations (number per unit volume) are proportional to

$$e^{-(\text{Energy}/kT)} \quad (2.1.1)$$

where $k=1.38 \times 10^{-23}$ J/K is the Boltzmann constant and T is the absolute temperature in Kelvin. Equation (2.1.1) is one of the fundamental principles of statistical mechanics: the concentration varies exponentially with the negative of the energy divided by the thermal energy kT .

It follows that the electron and hole densities in equilibrium, designated here by the symbols n and p , respectively, are related to the electrostatic potential ϕ by

$$\frac{p(\phi_1)}{p(\phi_2)} = e^{\frac{q(\phi_1 - \phi_2)}{kT}} \quad (2.1.2)$$

$$\frac{n(\phi_1)}{n(\phi_2)} = e^{\frac{q(\phi_1 - \phi_2)}{kT}} \quad (2.1.3)$$

where $q=1.6 \times 10^{-19}$ C is the electronic charge. Because of their negative charge, electrons are attracted to regions of higher electric potential. The opposite is true for holes. From equations (2.1.2) and (2.1.3) it follows that, in thermal equilibrium, the pn product is constant. This result, valid for both homogeneous and inhomogeneous material, is a specific case of the mass-action law [2].

Calling n_i the concentration of electrons (and holes) in an intrinsic semiconductor, the mass-action law is written as [1]

$$np = n_i^2. \quad (2.1.4)$$

For MOS compact models the case of a homogeneous substrate (constant doping concentration) is the most important. Calling n_0 and p_0 the equilibrium electron and hole concentrations, respectively, deep in the bulk of a homogeneous semiconductor, where charge neutrality holds, and choosing the potential reference $\phi=0$ in the neutral bulk, it follows that

$$p = p_0 e^{\frac{q\phi}{kT}} = p_0 e^{-u} \quad (2.1.5)$$

$$n = n_0 e^{\frac{q\phi}{kT}} = n_0 e^u \quad (2.1.6)$$

where $u = \phi / \phi_t$ is the normalized electrostatic voltage and $\phi_t = kT / q$ is the thermal voltage.

The charge density inside the semiconductor results from an imbalance between positive and negative charges. Four types of charges, electrons, holes, ionized acceptors (negatively charged) and ionized donors (positively charged) must be considered (Appendix B). Thus, the charge density ρ is given by

$$\rho = q(p - n + N_D - N_A) \quad (2.1.7)$$

where N_D and N_A are the ionized donor and acceptor densities, respectively.

Deep in the bulk of a uniformly doped semiconductor, where charge neutrality holds, the carrier concentrations at equilibrium are obtained from charge neutrality

$$p - n + N_D - N_A = 0 \quad (2.1.8)$$

and the mass-action law described in equation (2.1.4).

2.2 The field effect in bulk semiconductors

Field effect is the name given to the experiment in which an electric field is applied to the surface of a semiconductor. In the case of a metal the electric field does not penetrate at all in static conditions, while in the case of an ideal insulator the electric field lines pass through it. A semiconductor is an intermediate and important case, in which there is some penetration of the field into the material under static conditions leading to a modification in the conductance of the sample.

2.2.1 Fundamental considerations

The applied field should vary sufficiently slowly, so that the semiconductor is in thermal equilibrium. In addition, an idealized field effect model considers that the semiconductor doping is uniform and that the properties of the semiconductor are valid from the surface of the

semiconductor, *i.e.*, electrons and holes occur only as free carriers and their densities are given by Eqs. (2.1.5) and (2.1.6). In fact, the effect of the surface is to create allowed energy states for the electrons in the proximity of the surface. The term trap is applied to these states to indicate that they are electronic bound states, *i.e.*, electrons occupying these states are not free to move. Interface traps are a fundamental nonideality of a semiconductor surface (interface). For a high quality Si-SiO₂ interface, the interface trap density is so low that its effect can be disregarded and the idealized model presented in this section is appropriate.

The effect of interface traps will be considered later, in section 2.5, because it is sometimes important, *e.g.*, for stressed MOS transistors.

2.2.2 Poisson-Boltzmann equation in bulk semiconductors [3], [4]

Compact models for field effect transistors are based on the decomposition of the three (or two)-dimensional problem into two one-dimensional problems. Concerning the field effect, the applied field is assumed to be perpendicular to the surface so that the Poisson equation can be considered only in that direction.

Let x stand for the distance along this direction, with $x=0$ at the surface, and x going positive inwards as shown in Fig. 2.1. The one-dimensional Poisson equation (Appendix A) relating the electrostatic potential ϕ to the electric charge density ρ is

$$\frac{d^2\phi}{dx^2} = -\frac{dF}{dx} = -\frac{\rho}{\epsilon_s} \quad (2.2.1)$$

where F is the electric field and ϵ_s is the permittivity of the material.

Substituting the charge density ρ from expression (2.1.7) into (2.2.1) gives

$$\frac{dF}{dx} = \frac{q}{\epsilon_s} (p - n + N_D - N_A). \quad (2.2.2)$$

For bulk technologies, the semiconductor substrate is so thick that it can be considered of infinite depth for modeling purposes. Deep in the

bulk, charge neutrality holds, and the carrier densities n_0 and p_0 verify that

$$p_0 - n_0 + N_D - N_A = 0. \quad (2.2.3)$$

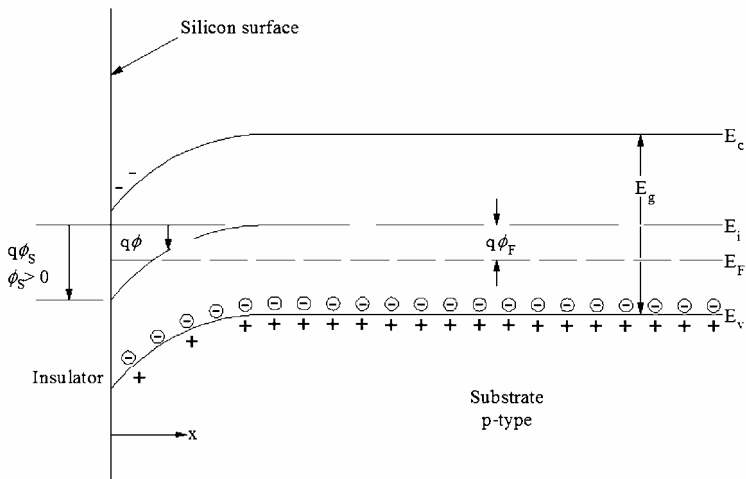


Fig. 2.1 Energy band diagram in a p-type semiconductor. The potential ϕ is defined as zero in the bulk. ϕ_s is the surface potential and ϕ_F is the equilibrium Fermi level (see equation (2.3.27)). The figure shows the case of positive (inwards) applied field. The maximum electron concentration occurs at the silicon surface, where the electron energy is minimal. The opposite holds for holes. The negative stored charge in the semiconductor results from both the excess of electrons and the lack of holes. (After [3].)

Substituting (2.1.5), (2.1.6), and (2.2.3) into (2.2.2) we obtain the Poisson-Boltzmann equation for the potential

$$\frac{d^2\phi}{dx^2} = -\frac{dF}{dx} = -\frac{q}{\epsilon_s} (p_0 e^{-u} - n_0 e^u + n_0 - p_0). \quad (2.2.4)$$

The Poisson-Boltzmann equation needs to satisfy the following boundary conditions

$$\phi = 0, \quad \frac{d\phi}{dx} = 0 \quad \text{for } x = \infty. \quad (2.2.5)$$

Deep in the interior of the material the potential approaches 0 V and the electric field vanishes. In general, it is not possible to integrate the

Poisson-Boltzmann equation twice using elementary functions. In particular cases, *e.g.* for an intrinsic semiconductor, an explicit expression for the potential profile $\phi(x)$ in terms of elementary functions is available [5], [6]. As a consequence, the analysis of the field effect is based on the first integral of (2.2.4) as we will show in the following.

Using now the definition of the one-dimensional electric field, it directly follows that $dx = -d\phi / F$. Differential equation (2.2.4) can thus be written as

$$\frac{1}{2\phi_t} \frac{dF^2}{du} = -\frac{q}{\epsilon_s} (p_0 e^{-u} - n_0 e^u + n_0 - p_0). \quad (2.2.6)$$

Considering the boundary conditions (2.2.5) at $x=\infty$ ($u=0$) for the potential and the field, and calculating the definite integral of the right-hand side of (2.2.6) we obtain

$$F^2 = \frac{2q\phi_t}{\epsilon_s} [p_0 (e^{-u} + u - 1) + n_0 (e^u - u - 1)]. \quad (2.2.7)$$

Equation (2.2.7) gives the square of the electric field $F = -d\phi/dx$ in terms of the normalized local potential. For a positive field going into the semiconductor, the potential decreases in the inward direction, and because the origin of the potential is at infinity, the potential is positive inside the semiconductor. Consequently, to calculate the electric field, the meaningful sign of the square root corresponds to $F>0$ when $u>0$ and $F<0$ when $u<0$.

Eq. (2.2.7) is very general, and is valid for p-type, n-type and intrinsic semiconductors. A simple interpretation of (2.2.7) is based on the behavior of non-negative symmetric functions $(e^{-u} + u - 1)$ and $(e^u - u - 1)$ (Fig. 2.2). As expected, the squared electric field is never negative and goes to zero for $\phi=0$. The term in the right-hand side preceded by the factor p_0 corresponds to the contribution of holes to the electric field and that preceded by the factor n_0 corresponds to the contribution of electrons. Increasing positive values of the potential increase the electron concentration and decrease the hole concentration.

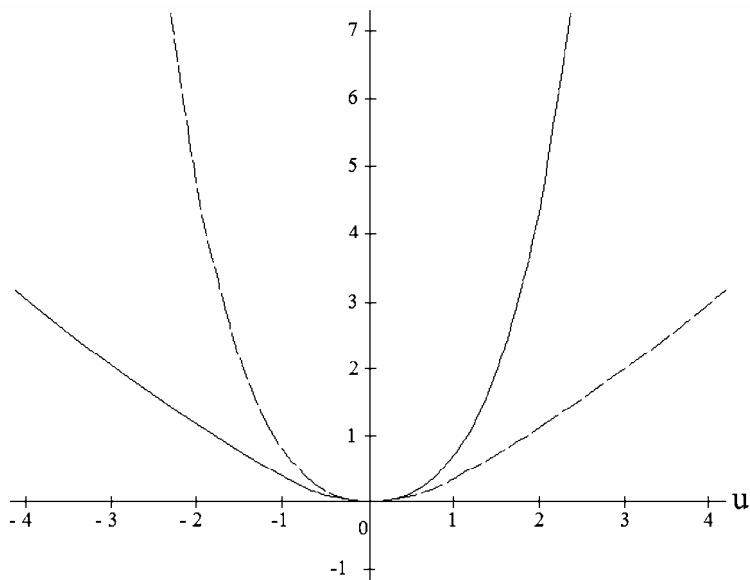


Fig. 2.2 Plot of functions (-----) $(e^{-u} + u - 1)$ and (—) $(e^u - u - 1)$.

Let us now consider the specific case of a p-type semiconductor ($p_0 \gg n_0$) that will be used throughout this book to develop the n-channel transistor model. The value of the electric field results from the contribution of the hole and electron terms, $p_0(\exp(-u) + u - 1)$ and $n_0(\exp(u) - u - 1)$, respectively. For negative values of the potential ($u < 0$), the contribution of holes to the charge is predominant, and the semiconductor is in accumulation, $F \sim e^{-u/2}$. For positive and small values of the potential $F \sim u^{1/2}$, because the predominant contribution to the space charge is the depletion of holes ($p_0 \gg n_0$); in this case, the semiconductor is said to be in depletion or weak inversion. Finally, for values of the potential such that $n_0 e^u > p_0 u$, the semiconductor is in strong inversion and $F \sim e^{u/2}$. Fig. 2.3 shows a plot of the field function F against the potential.

The traditional approach to transistor modeling has been to develop separate models for each region of operation. Up to four regions have been considered, accumulation, and weak, moderate and strong inversion. In this book we emphasize next generation models that avoid modeling each of the operation regimes with a specific equation.

Consequently, we will not emphasize specific approximations of (2.2.7) for each operation regime.

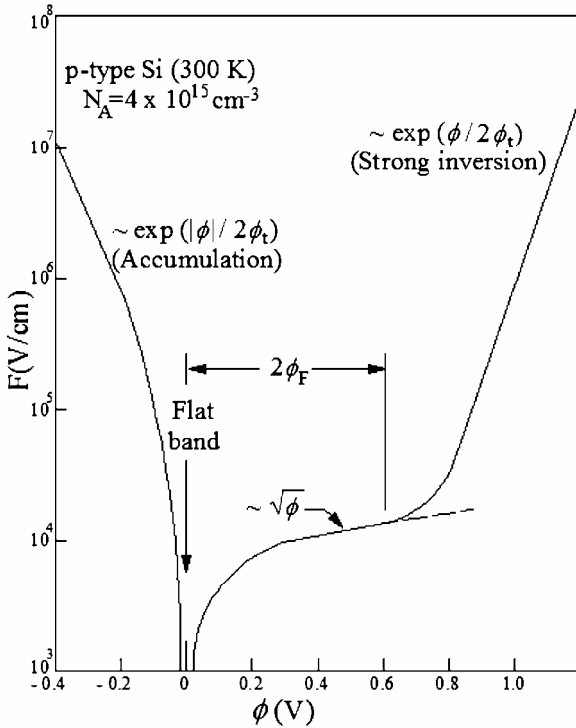


Fig. 2.3 Electric field in the semiconductor as a function of the potential ϕ for p-type silicon. (After [7].)

2.3 The ideal two-terminal MOS structure [8]

The MOS structure, consisting of a metal-oxide (SiO_2)-semiconductor (Si) sandwich, made the fabrication of the first practical surface field effect transistor possible in 1960, and until now it is the core structure of Very Large Scale Integration (VLSI) systems.

The MOS structure (Fig. 2.4) functions as a two-terminal capacitor in which a metal plate is separated from the semiconductor substrate by a thin insulator of a thermally grown SiO_2 layer. The unique properties of the SiO_2 -Si system are the reduced amount of interface traps and the high

dielectric strength of the SiO_2 [9]. The most common metal plate materials are aluminum and heavily doped polycrystalline silicon (polysilicon or poly). A second metal layer along the back provides an electrical contact with the silicon substrate.

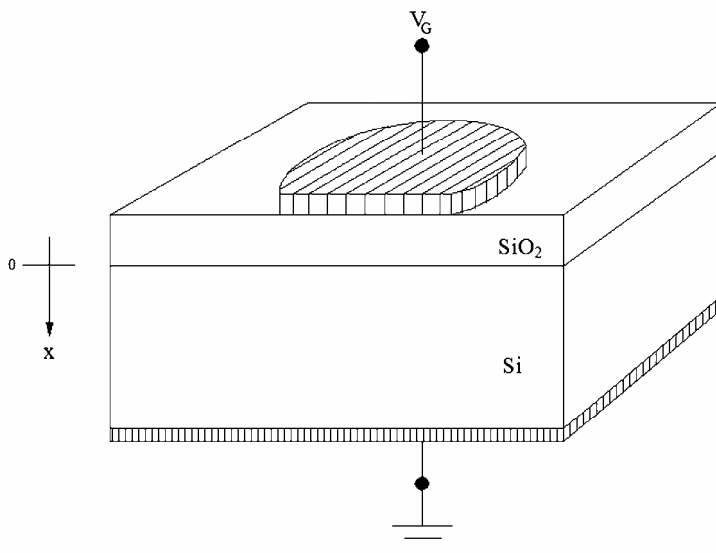


Fig. 2.4 The metal-oxide-semiconductor capacitor.

The ideal MOS structure has the following properties [10]: (1) the metal gate is equipotential; (2) the oxide is a perfect insulator with zero transport current and no charge inside it or at its interfaces; (3) there are no interface traps at the SiO_2 -Si interface; (4) the semiconductor is uniformly doped; (5) the semiconductor is sufficiently thick so that a field free region (“bulk”) exists far from the interface; (6) the back gate contact is ideal (ohmic); (7) the structure is one-dimensional, *i.e.*, the electric field lines are perpendicular to the surface; and (8) the potential contact between the metal and semiconductor is zero. The analysis of the ideal MOS structure is of great use because all the idealized properties are approached in many real MOS structures, except for property (8). However, the inclusion of the potential contact value in the model requires a simple voltage shift of the capacitive characteristics. The

inclusion of other nonidealities, such as the polydepletion effect, quantum effects, and interface traps, will be presented in section 2.5.

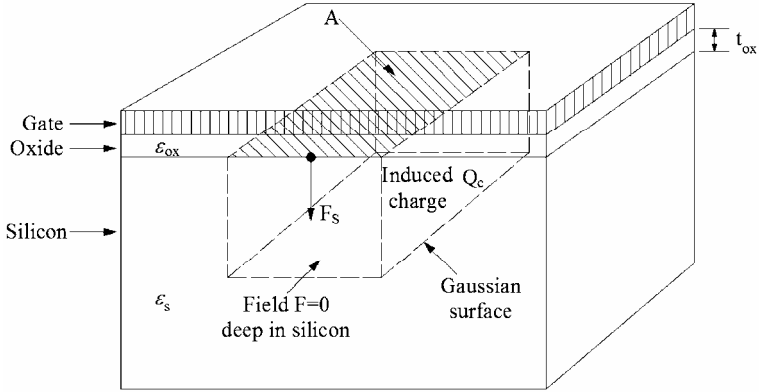


Fig. 2.5 Illustration of a Gaussian surface enclosing the charge Q_C stored in the semiconductor. Gauss' law (Appendix A) gives the relationship between the enclosed charge and the field F_s in the semiconductor interface.

For the ideal MOS structure with zero contact potential between the metal and semiconductor, any applied gate-to-substrate voltage V_G appears partly across the oxide and partly across the semiconductor. Thus

$$V_G = \phi_{ox} + \phi_s, \quad (2.3.1)$$

where ϕ_{ox} and ϕ_s are the potential drops across the oxide and the semiconductor, respectively. For an ideal insulator (with no charge inside) the electric field is given by

$$F_{ox} = \frac{\phi_{ox}}{t_{ox}}, \quad (2.3.2)$$

where t_{ox} is the oxide thickness. Assuming the oxide semiconductor interface to be ideal (with no interface charges), the continuity of the normal component of the electric displacement gives (Appendix A)

$$\epsilon_{ox} F_{ox} = \epsilon_s F_s. \quad (2.3.3)$$

Finally, the electric field at the semiconductor interface is given by Gauss' law (Fig. 2.5) as

$$F_s = -\frac{Q'_C}{\epsilon_s} \quad (2.3.4)$$

where Q'_C is the total semiconductor charge per unit area. The minus sign in (2.3.4) holds because the positive orientation of the field is in the inward direction of the Gaussian surface.

The substitution of (2.3.4) into (2.3.3) results in an expression that, substituted into (2.3.2), gives the voltage drop across the oxide as

$$\phi_{ox} = -\frac{t_{ox}}{\epsilon_{ox}} Q'_C = -\frac{Q'_C}{C'_{ox}} \quad (2.3.5)$$

where $C'_{ox} = \epsilon_{ox} / t_{ox}$ is the oxide capacitance per unit area.

For the ideal MOS capacitor, the fundamental expression that relates the applied voltage V_G with the surface potential ϕ_s and the semiconductor space charge Q'_C

$$V_G = \phi_s - \frac{Q'_C}{C'_{ox}} \quad (2.3.6)$$

results from the substitution of (2.3.5) into (2.3.1).

2.3.1 Small signal equivalent capacitive circuit

Because the MOS structure stores charge, its electrical behavior is capacitive. Due to the highly non linear charge-voltage characteristic of the MOS structure, the small signal capacitance as a function of the dc voltage is the commonly measured electrical characteristic of the MOS capacitor. For the ideal MOS, the insulator blocks dc current and, neglecting losses, the parallel conductance is zero. Let's start with the low-frequency or quasi-static characteristic, where the frequency is so low that the time-varying ac state can be considered as a succession of equilibrium (dc) states in the semiconductor.

If Q'_G is the gate charge per unit area, then $Q'_G = -Q'_C$, and the small-signal capacitance per unit area is given by

$$C'_{gb} = \frac{dQ'_G}{dV_G} = -\frac{dQ'_C}{dV_G} \quad (2.3.7)$$

The substitution of (2.3.6) into (2.3.7) yields

$$C'_{gb} = -\frac{dQ'_C}{d\phi_s - \frac{dQ'_C}{C'_{ox}}} = \frac{1}{-\frac{d\phi_s}{dQ'_C} + \frac{1}{C'_{ox}}} \quad (2.3.8)$$

or

$$C'_{gb} = \frac{1}{\frac{1}{C'_c} + \frac{1}{C'_{ox}}} \quad (2.3.9)$$

where $C'_c = -dQ'_C/d\phi_s$ is the semiconductor (space-charge) capacitance per unit area. Thus, the capacitance of the MOS structure is the series combination of the oxide and the semiconductor capacitances. The equivalent circuit for the ideal MOS capacitor is given in Fig. 2.6 (a).

To calculate the semiconductor capacitance, the total charge per unit area in the semiconductor is obtained from Gauss' law (Fig. 2.5) and expression (2.2.7) for the electric field F as

$$-Q'_C = \epsilon_s F_s = \text{sign}(u_s) \sqrt{2q\phi_i \epsilon_s} \sqrt{p_0(e^{-u_s} + u_s - 1) + n_0(e^{u_s} - u_s - 1)} \quad (2.3.10)$$

where u_s is the normalized surface potential and F_s is the electric field at the semiconductor interface.

Differentiating (2.3.10) with respect to the surface potential ϕ_s gives the expression for the semiconductor capacitance C'_c

$$C'_c = \frac{q}{F_s} \left[n_0(e^{u_s} - 1) + p_0(1 - e^{-u_s}) \right] = \frac{q}{F_s} (n_s - n_0 + p_0 - p_s) \quad (2.3.11)$$

where n_s and p_s are the electron and hole concentrations at the interface, respectively. The semiconductor capacitance can be separated into the electron and hole storage capacitances (see Fig. 2.6 (b)), given by the terms containing the electron and hole concentrations n and p , respectively. This separation has a clear physical meaning as can be

verified in the calculation of the total stored electron and hole charges [2]. The total charge density in the semiconductor is

$$Q'_C = \int_0^{\infty} \rho dx = q \left[\int_0^{\infty} (p - p_0) dx + \int_0^{\infty} (n_0 - n) dx \right] = Q'_P + Q'_N. \quad (2.3.12)$$

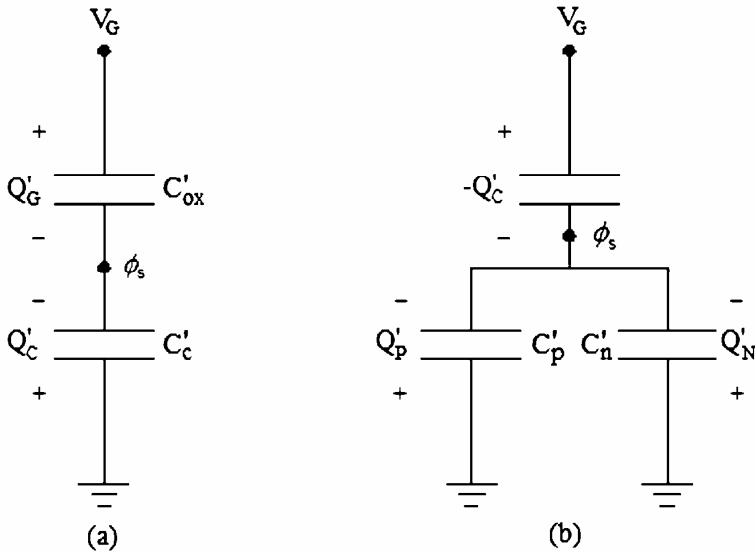


Fig. 2.6 Equivalent circuits for the MOS capacitor: (a) Two-capacitance model; (b) Splitting of the semiconductor capacitance into electron and hole storage capacitances.

Changing the integration variable from length x to potential ϕ gives

$$Q'_C = Q'_P + Q'_N = q \int_0^{\phi_s} \frac{(p - p_0)}{F} d\phi + q \int_0^{\phi_s} \frac{(n_0 - n)}{F} d\phi. \quad (2.3.13)$$

Consequently, as expected, the hole and electron storage capacitance contributions to the total semiconductor capacitance are

$$\begin{aligned} C'_p &= -\frac{dQ'_P}{d\phi_s} = \frac{q}{F_s} (p_0 - p_s) \\ C'_n &= -\frac{dQ'_N}{d\phi_s} = \frac{q}{F_s} (n_s - n_0). \end{aligned} \quad (2.3.14)$$

The theoretical small-signal capacitance vs. voltage (C-V) curve can be calculated as follows. Substituting (2.3.10) into (2.3.6) we obtain the equation to determine the surface potential ϕ_s in terms of the applied bias V_G . This nonlinear equation has to be numerically solved with high accuracy due to the exponential dependence of the carrier concentrations on the potential. A typical ϕ_s vs. V_G characteristic is seen in Fig. 2.7.

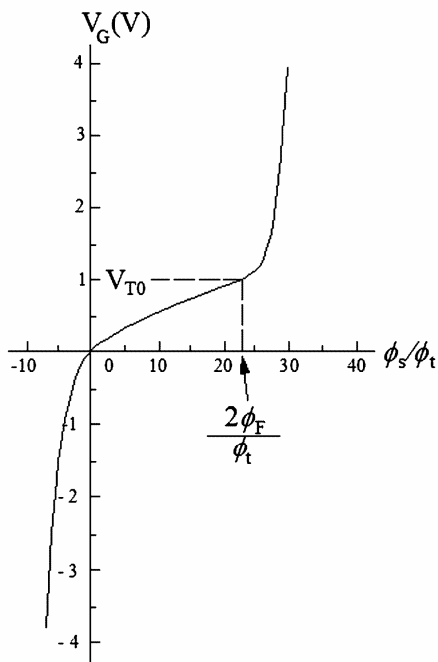


Fig. 2.7 Relationship between surface potential (semiconductor band-bending) ϕ_s and applied gate voltage V_G for the MOS capacitor. (After [10].) V_{T0} is the threshold voltage (see Eq. (2.4.52)).

Deep in accumulation and inversion, the surface potential presents a reduced (logarithmic) variation with the gate voltage due to the exponential dependence of the dominant carrier charge on the surface potential. In accumulation, ϕ_s is usually above $-8\phi_t$, -200mV at room temperature. In inversion, ϕ_s is usually below 1V.

Once the surface potential is (accurately) known, all the other variables are easily determined. Expressions (2.3.11) and (2.3.10) allow the calculation of the semiconductor capacitance in terms of the surface potential. The total gate capacitance is calculated from (2.3.9). The problem is numerically well-conditioned except under the flat-band condition $\phi_s = 0$, $F_s = 0$. The first order series expansion around $\phi_s = 0$ for the electric field, given by equation (2.3.10), is

$$F_s \approx \sqrt{\frac{q\phi_t}{\epsilon_s}(p_0 + n_0)}u_s. \quad (2.3.15)$$

For a p-type semiconductor, $p_0 \gg n_0$. Noting that, for $u_s \rightarrow 0$ we have $p_0 - p_s = p_0(1 - \exp(-u_s)) \approx p_0 u_s$ and $n_s - n_0 \approx n_0 u_s$, the substitution of (2.3.15) into (2.3.14) gives

$$C'_{pfb} = \sqrt{\frac{q\epsilon_s p_0}{\phi_t}} = \sqrt{\frac{q^2 \epsilon_s p_0}{kT}}; \quad C'_{nfb} = C'_{pfb} \left(\frac{n_0}{p_0} \right) \quad (2.3.16)$$

The flat-band capacitance is essentially due to the storage of majority carriers (holes in this case). The flat-band capacitance can also be written as

$$C'_{fb} \approx C'_{pfb} = \frac{\epsilon_s}{L_D} \quad (2.3.17)$$

where L_D is the extrinsic Debye length (see Appendix B).

2.3.2 Ideal C-V plots

A plot of the theoretical MOS C-V curve calculated according to the procedure of the previous section is shown in Fig. 2.8. To develop an insightful understanding of the shape of the C-V curve, let us consider the case of a p-type semiconductor. When the gate voltage is negative, the semiconductor is in accumulation, $\exp(-u_s) \gg u_s - 1$ and $p_0 \exp(-u_s) \gg n_0(\exp(u_s) - u_s - 1)$; consequently

$$-Q'_C \approx -\sqrt{2q\phi_t \epsilon_s p_0} e^{-u_s/2} \quad (2.3.18)$$

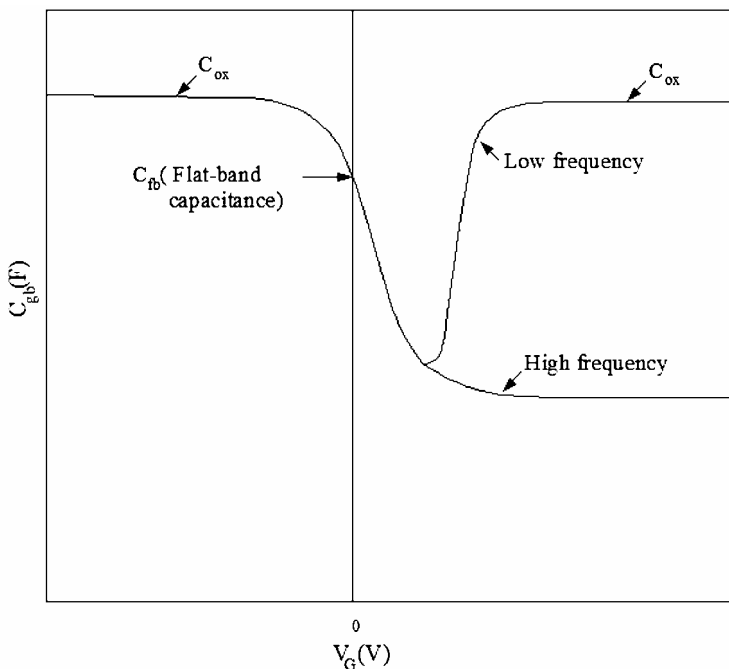


Fig. 2.8 MOS C-V curve. (After [7].)

and

$$C'_c = \frac{Q'_C}{2\phi_t} = -\frac{C'_{ox}(V_G - \phi_s)}{2\phi_t}. \quad (2.3.19)$$

In (2.3.19), Q'_C was written in terms of the potential using (2.3.6). The gate capacitance obtained as the series connection of C'_{ox} and C'_c is

$$\frac{1}{C'_{gb}} = \frac{1}{C'_{ox}} \left(1 + \frac{2\phi_t}{|V_G - \phi_s|} \right). \quad (2.3.20)$$

Expression (2.3.20) gives the asymptotic behavior of the MOS capacitance in accumulation. Deep in accumulation $|V_G - \phi_s| \gg 2\phi_t$, the MOS capacitance approaches the oxide capacitance because the holes concentrate at the interface. Considering a typical absolute value of $-\phi_s$ below 200mV, the gate capacitance approaches the oxide capacitance for

gate voltages under -1V. The gate capacitance in accumulation is nonlinear and (2.3.20) is a useful design equation to calculate the distortion introduced by an MOS capacitor in accumulation in a linear circuit [11]. A default value of $\phi_s = 0$ can be used in (2.3.20) to assess distortion. For improved accuracy, the value of ϕ_s calculated for the bias point can be employed. (2.3.20) can also be used to predict the capacitance of an MOS capacitor employed as a varactor. Expression (2.3.20) is meaningless at flat-band, but we have already calculated the semiconductor capacitance for this case (see Eq. (2.3.16)).

Consider now that ϕ_s is positive and greater than several ϕ_t , but not too large, so that the following inequalities characterizing depletion, $\exp(-u_s) \ll u_s - 1$ and $p_0(u_s - 1) \gg n_0(\exp(u_s) - u_s - 1)$, hold. Neglecting ϕ_t in comparison with ϕ_s we rewrite (2.3.10) as

$$Q'_C \approx -\sqrt{2q\epsilon_s p_0 \phi_s}. \quad (2.3.21)$$

Consequently, the semiconductor capacitance is given by

$$C'_c = \frac{\sqrt{2q\epsilon_s p_0}}{2\sqrt{\phi_s}} = \frac{\gamma C'_{ox}}{2\sqrt{\phi_s}} \quad (2.3.22)$$

where

$$\gamma = \frac{\sqrt{2q\epsilon_s p_0}}{C'_{ox}} \quad (2.3.23)$$

is the body effect factor. Substituting (2.3.21) into (2.3.6) we obtain

$$\sqrt{\phi_s} = \sqrt{\frac{\gamma^2}{4} + V_G} - \frac{\gamma}{2}. \quad (2.3.24)$$

Finally, substituting (2.3.24) into (2.3.22) and calculating the total MOS capacitance through (2.3.9) we obtain

$$C'_{gb} = \frac{C'_{ox}}{\sqrt{1 + \frac{4V_G}{\gamma^2}}}. \quad (2.3.25)$$

When the semiconductor operates in depletion, the capacitance decreases with the square root of the gate voltage. Equation (2.3.25) has

a limited range of validity. At flat-band, (2.3.25) gives a gate capacitance equal to the oxide capacitance, corresponding to an infinite value of the flat-band capacitance. The other limit of validity of (2.3.25) is for the semiconductor entering inversion, when the capacitance begins to increase due to the inversion layer near the semiconductor surface. Assuming that the inversion charge is dominant, an expression similar to that of accumulation is obtained [11], [12]. The strong inversion capacitance is also given by (2.3.20), but here the absolute value of the surface potential is higher than in accumulation. Even so, for values of the gate voltage above 2V the MOS capacitance approaches the oxide capacitance; the physical reason for this behavior is that, as the gate voltage changes, the electron charge varies in close proximity to the Si-SiO₂ interface, thus screening the depletion charge variations. As in accumulation, the gate capacitance is nonlinear and (2.3.20) is a useful design equation to calculate either the distortion introduced by an MOS capacitor in inversion or the voltage-dependent capacitance of an MOS varactor.

The C-V curve gives direct information about the penetration of the field in the semiconductor (minimum in accumulation, maximum in depletion, etc.) and about the varying charges (majority and/or minority carriers) allowing a direct assessment of the field effect. C-V characteristics are widely used to characterize the semiconductor, oxide, and Si-SiO₂ interface [13], [14]. Test capacitors are customarily measured to monitor process lines, and a vast literature exists concerning the MOS capacitor as a technology characterization tool (see, for example, references in [14]). For technology characterization purposes, the inclusion of accurate models of the different non-idealities is paramount. To develop compact models of the MOS transistor we must keep the capacitive model as simple as possible. An example of this simplicity is the traditional strong inversion model, which is based on the assumption of a constant gate capacitance, equal to the oxide capacitance.

The quasi-static model of the MOS capacitance is applicable if the carriers can follow the applied signal. For majority carriers, this is the case for frequencies as high as tens or even hundreds of MHz. For minority carriers, the response time depends on the generation rate of the

material and can be very long, particularly for very pure silicon samples, because the generation rate depends on the existing defects. Consequently, two limit C-V curves can be determined for MOS capacitors (Fig. 2.8).

In some cases, the low-frequency curve must be determined for very low signal frequency, sometimes under 10 Hz. A typical frequency value for the high-frequency curve is 1MHz. The model of the high-frequency curve is easily obtained disregarding the minority carrier contribution to the capacitance, *i.e.*, considering that the minority carrier number is fixed by the dc bias and the majority carriers follow the applied signal; consequently, the MOS depletion behavior extends into the inversion region, and the curve does not present a minimum, as seen in Fig. 2.9.

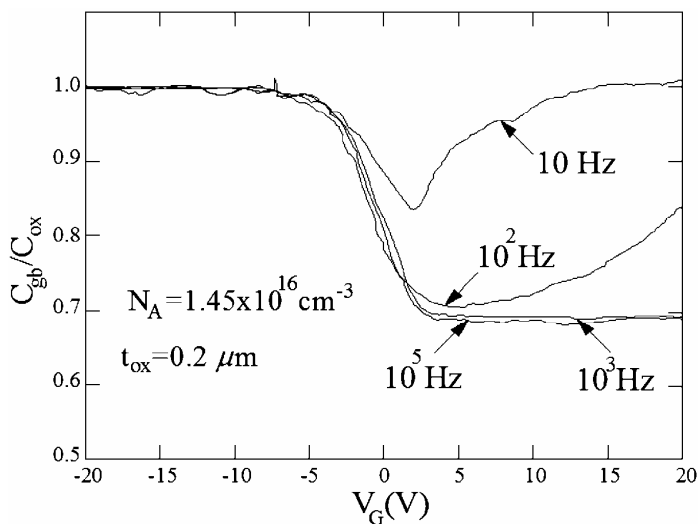


Fig. 2.9 The effect of measurement frequency on the capacitance-voltage characteristics of an MOS capacitor. (After [15].)

For MOS transistor modeling the relevant C-V curve is the low-frequency one. In effect, minority carriers under the gate of the transistor follow the high frequency variation of the gate voltage because the source and drain regions are able to supply very quickly the necessary

number of minority carriers and, consequently, the response time is independent of the generation rate in the bulk.

2.3.3 Effect of work function difference

The energy of carriers inside a (semi)conductor depends on the material. When two different materials J_1 and J_2 are placed in (ideal) contact, carriers tend to flow to the material where their energy is lower.

As carriers cross the junction, a charge imbalance between the two materials occurs, which in turn generates an electric field that counteracts the carrier flow. Eventually, when equilibrium is reached, an electrostatic potential difference $\phi_{J1,J2}$ is established between the two materials, called the potential contact (see Fig. 2.10), so that there is no more net carrier flow [16]. Conventionally, the contact potentials are expressed in terms of the work functions of the two materials.

The work function W of a material is the energy necessary to remove an electron from the Fermi energy level to an infinite distance from the sample. The contact potential in terms of the work functions is

$$\phi_{J1,J2} = -\frac{(W_{J1} - W_{J2})}{q}. \quad (2.3.26)$$

A straightforward definition of the Fermi potential of a semiconductor taken from the concept of contact potential is illustrated in Fig. 2.11 [16]: The Fermi potential of a homogeneous extrinsic semiconductor can be defined as the contact potential between the intrinsic semiconductor and the extrinsic semiconductor. The contact potential appears between the neutral extrinsic and intrinsic parts, where the electron (hole) densities are n_0 (p_0) and n_i , respectively. It follows from (2.1.2) and (2.1.3) that

$$-\phi_J = \phi_F = \frac{kT}{q} \ln \left(\frac{p_0}{n_i} \right) = -\frac{kT}{q} \ln \left(\frac{n_0}{n_i} \right). \quad (2.3.27)$$

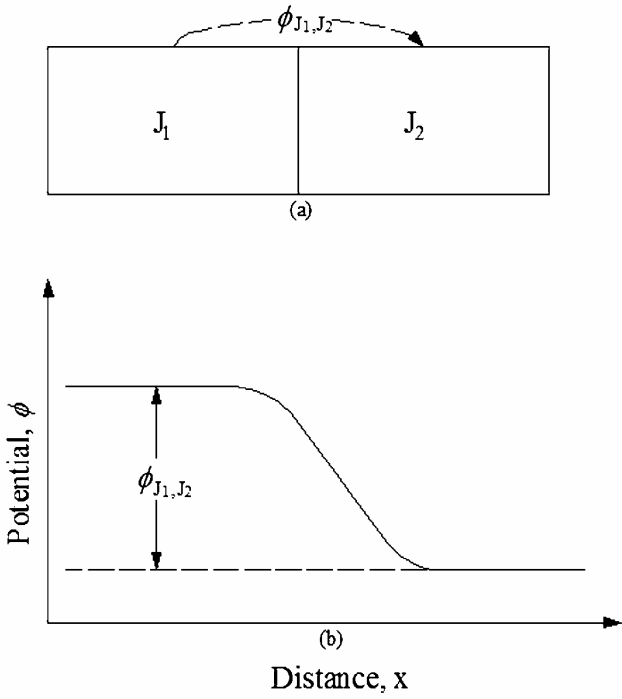


Fig. 2.10 (a) Two different materials in contact (either can be a metal or a semiconductor); (b) potential vs. distance. (After [16].)

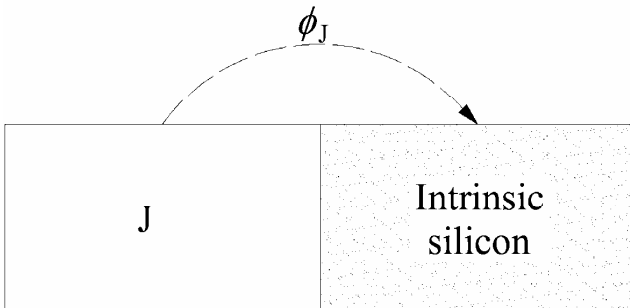


Fig. 2.11 Junction of a material J and intrinsic silicon; definition of the potential ϕ_J . (After [16].)

For a p-type semiconductor the Fermi potential is positive and for an n-type material it is negative.

Intrinsic silicon is used here only as a reference material; the potential difference between two conductors (semiconductors) is the difference between their contact potentials with respect to intrinsic silicon. For several materials in series (assumed to be at the same temperature), as depicted in Fig. 2.12, the electrostatic potential difference $\phi_{J1,Jn}$ depends only on the first and last material [16]. In effect,

$$\phi_{J1,Jn} = (\phi_{J1} - \phi_{J2}) + (\phi_{J2} - \phi_{J3}) + \dots (\phi_{Jn-1} - \phi_{Jn}) = \phi_{J1} - \phi_{Jn}. \quad (2.3.28)$$

For a homogeneous semiconductor in equilibrium, writing the electron and hole densities n_0 and p_0 deep in the bulk in terms of the Fermi potential, and normalizing all potentials with respect to the thermal potential $\phi_t = kT/q$, (2.1.5) and (2.1.6) can be rewritten as

$$p = n_i e^{u_F - u} \quad (2.3.29)$$

$$n = n_i e^{u - u_F} \quad (2.3.30)$$

where $u_F = \phi_F / \phi_t$ is the normalized Fermi potential.

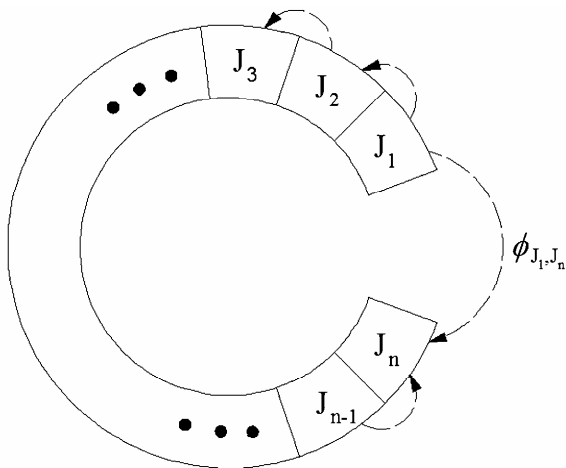


Fig. 2.12 Several materials in series. (After [16].)

2.3.4 The flat-band voltage

In equilibrium (with the two terminals shortened/open) the contact potential between the gate and the semiconductor substrate of the MOS structure depends only on the gate and substrate materials, being independent of any material in-between (see equation (2.3.28)). This contact potential is given by the difference in the work functions of the gate and substrate materials, as indicated in (2.3.26). For a nonzero contact potential, charges are induced in the gate and the semiconductor for $V_{gb} = 0$.

By applying an external gate-to-bulk voltage opposite to the gate-bulk contact potential, the effect of the work function difference is counterbalanced, and zero charge inside the semiconductor (flat-band condition) is obtained, as shown in Fig. 2.13. This voltage is called the flat-band voltage V_{FB} , because under this condition the semiconductor is equipotential; thus, the energy bands are flat.

The existence of charges inside the insulator and at the semiconductor-insulator interface also induces a semiconductor charge at zero bias. For high quality Si-SiO₂ interfaces this effect can be almost negligible [12], but sometimes, for stressed devices, for example, it must be considered. The effect of the oxide charges on the semiconductor depends on their position inside the oxide. However, it is not possible for the user of the device to determine the location of the charges from electrical measurements. Consequently, the effect of the oxide is represented in compact models by an effective oxide interface charge Q'_o . The potential drop across the oxide is

$$\phi_{ox} = -\frac{Q'_o}{C'_{ox}}. \quad (2.3.31)$$

The flat-band voltage considering the effects of both the work function difference and oxide charges (Fig. 2.14) is

$$V_{FB} = \phi_{MS} - \frac{Q'_o}{C'_{ox}} \quad (2.3.32)$$

where

$$\phi_{MS} = \phi_{\text{bulk_material}} - \phi_{\text{gate_material}}. \quad (2.3.33)$$

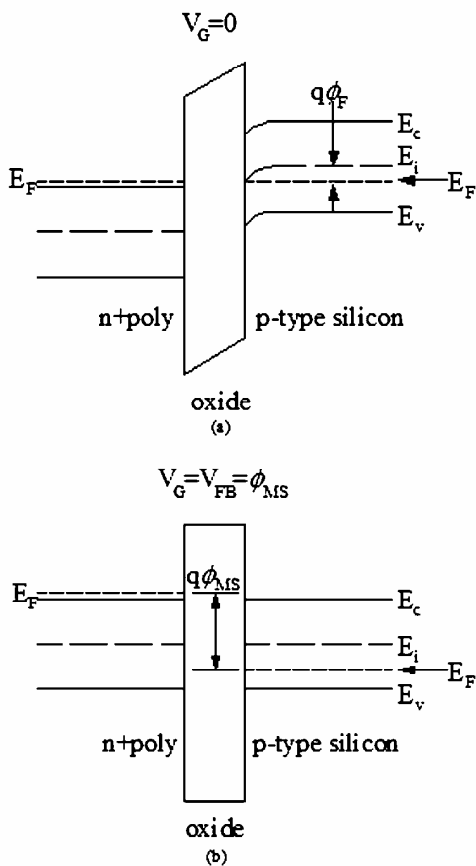


Fig. 2.13 Band diagram of an n^+ -polysilicon-gate p-type MOS capacitor biased at (a) zero gate voltage and (b) flat-band condition. (After [12].)

The effect of the oxide charges and potential contact between gate and bulk materials is to shift horizontally the C-V characteristic by an amount equal to V_{FB} , as can be seen in Fig. 2.15.

The inclusion of the flat-band voltage in equation (2.3.6) yields

$$V_G - V_{FB} = \phi_s - \frac{Q'_C}{C'_{ox}}, \quad (2.3.34)$$

which is known as the potential balance equation.

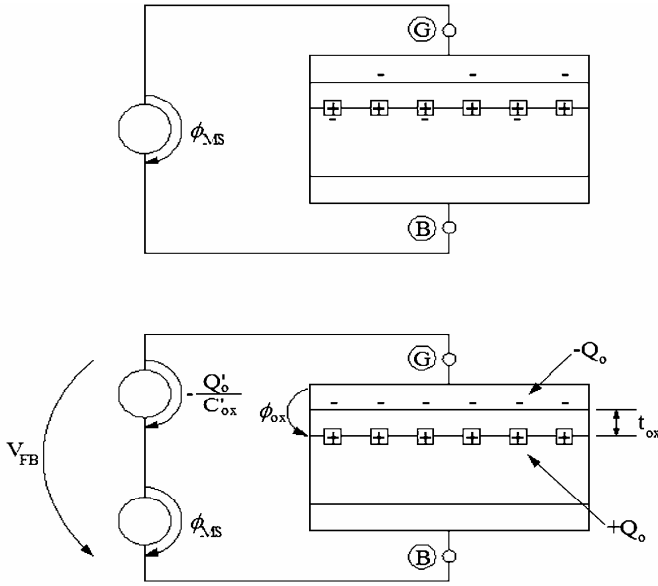


Fig. 2.14 Flat-band voltage definition for an MOS structure showing compensation of the metal-semiconductor work function only and compensation of both metal-semiconductor work function and equivalent oxide charge. (After [16].)

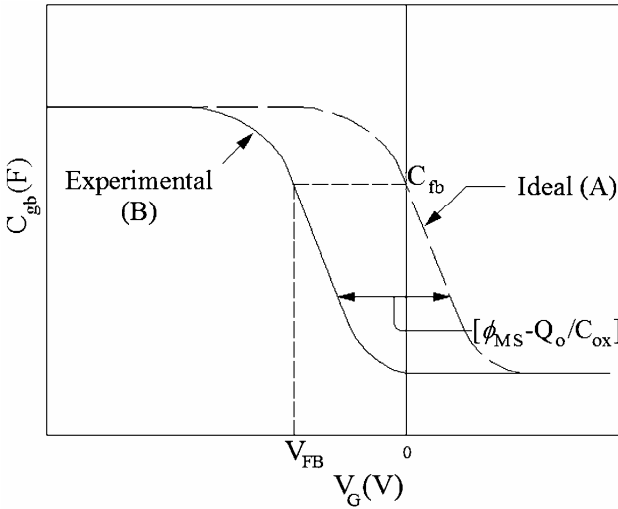


Fig. 2.15 Influence of the metal-semiconductor work function difference and oxide charges on the high-frequency C-V curve for an MOS capacitor. (After [42].)

2.4 The three-terminal MOS with uniformly doped substrate

The MOS transistor needs contact regions, necessarily of a type opposite to that of the substrate, to access the inversion channel. Thus, the modeling of the three-terminal structure or gate-controlled diode of Fig. 2.16 is an intermediate stage in the development of transistor models.

The analysis of the 3-terminal MOS structure combines the MOS capacitor and p-n junction theory, and is conveniently formulated in terms of the quasi-Fermi potentials, the subject of the next section.

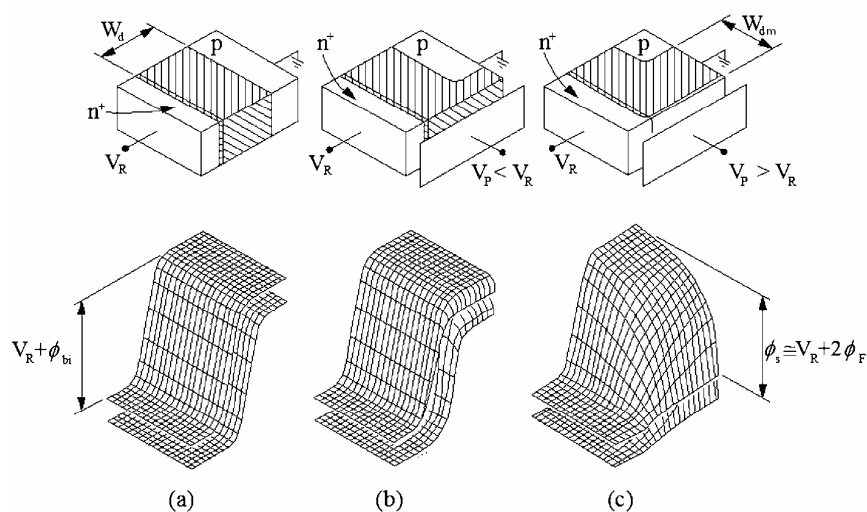


Fig. 2.16 Three terminal MOS structure on a p-type substrate and energy bands. The n^+ -p-substrate junction is reverse biased. The gate is biased at (a) flat-band, (b) depletion, and (c) inversion. (After [8].) V_P is the pinch-off voltage (see Eq. (2.4.49)) and $W_{d(m)}$ is the depletion width.

2.4.1 Quasi-equilibrium electron and hole concentrations

The quasi-Fermi potentials ϕ_p for holes and ϕ_n for electrons are defined in terms of carrier concentrations (p and n) and potential ϕ . In order to write the concentrations under non-equilibrium conditions, preserving the form of the Boltzmann equilibrium expressions, one simply replaces

the equilibrium Fermi potential with the quasi-Fermi potentials [17] as follows

$$p = n_i e^{u_p - u} \quad (2.4.1)$$

$$n = n_i e^{u - u_n} \quad (2.4.2)$$

or

$$\phi_p = \phi_t u_p = \phi + \phi_t \ln \left(\frac{p}{n_i} \right) \quad (2.4.3)$$

and

$$\phi_n = \phi_t u_n = \phi - \phi_t \ln \left(\frac{n}{n_i} \right). \quad (2.4.4)$$

In equilibrium, the quasi-Fermi potentials ϕ_n and ϕ_p reduce to the Fermi potential ϕ_F . In terms of the quasi-Fermi levels, the current densities are written in a very compact form. For electrons, the total current density in one dimension (y), including drift and diffusion, is given by (Appendix C)

$$J_n = qn\mu_n \left(-\frac{d\phi}{dy} \right) + qD_n \frac{dn}{dy}. \quad (2.4.5)$$

Using the Einstein relationship ($D_n = \mu_n \phi_t$) between the electron mobility μ_n and the diffusion coefficient D_n , (2.4.5) is rewritten as

$$J_n = -qn\mu_n \left(\frac{d\phi}{dy} - \frac{\phi_t}{n} \frac{dn}{dy} \right) = -qn\mu_n \frac{d\phi_n}{dy}. \quad (2.4.6)$$

Similarly, for holes

$$J_p = -qp\mu_p \left(\frac{d\phi}{dy} + \frac{\phi_t}{p} \frac{dp}{dy} \right) = -qp\mu_p \frac{d\phi_p}{dy}. \quad (2.4.7)$$

Let us now consider a biased p-n junction, as shown in Fig. 2.17. From the definition of quasi-Fermi levels in (2.4.1) and (2.4.2), the differences $(\phi_p - \phi)$ and $(\phi - \phi_n)$ in the neutral p and n regions, respectively, must be the same as in equilibrium, because the majority carrier densities are (almost) the same as in equilibrium. Given that the

applied reverse (forward) voltage appears as an increase (decrease) in the potential barrier inside the semiconductor, the quasi-Fermi levels must split by an amount equal to the applied reverse (forward) bias, to keep the majority carrier concentrations constant, *i.e.*,

$$\phi_n - \phi_p = V_C \quad (2.4.8)$$

as shown in Fig. 2.17. Note that $V_C > 0$ for a reverse biased junction.

Since the hole current is carried by a high hole concentration in the p-region and by a low hole concentration in the n-region, the gradient of ϕ_p must be much greater in the n- than in the p-region, as shown in Fig. 2.17. For the same reason, the variation in ϕ_n occurs mainly in the p-region. A summary of the properties of the p-n junction is given in Appendix E.

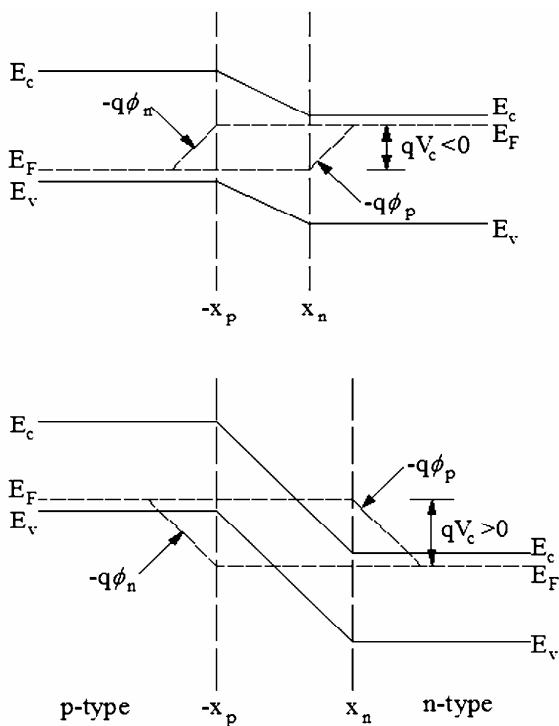


Fig. 2.17 Variation of the electron and hole quasi-Fermi levels for (a) forward-biased and (b) reverse-biased junction. (After [12].)

2.4.2 Surface potential master equation

The band diagram of the three-terminal MOS structure is shown in Fig. 2.18. To develop compact models, the quasi-Fermi levels over the space-charge region are considered constant along the x -direction (perpendicular to the surface) since there is no current flow along the x -direction. For the three-terminal MOS with the source-substrate junction reverse biased, as is the usual operating condition, we neglect the reverse current of the source-substrate junction, which gives a quite acceptable approximation.

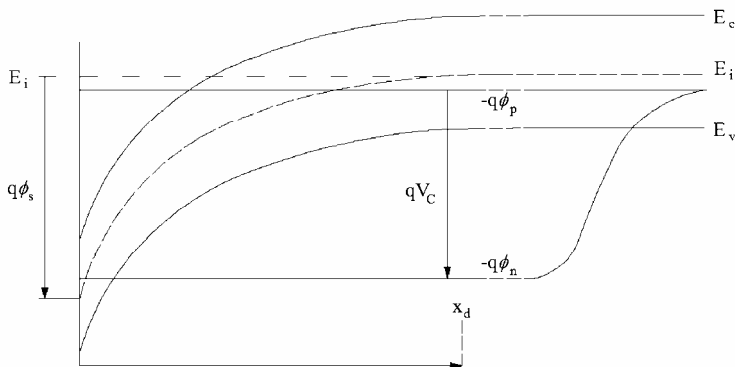


Fig. 2.18 The energy band diagram in the x -direction, perpendicular to the surface, for the three-terminal MOS structure. (After [5].)

The hole quasi-Fermi potential is essentially independent of the distance and coincides with the equilibrium Fermi level in the bulk. The constancy of the electron quasi-Fermi potential from the n^+ to the n -channel region represents the fact that the inverted n -channel is connected to the n^+ source region. In other words, the contact between the n^+ region and the channel implies the equilibrium of electrons in these two regions, the electron density varying according to the value of the electrostatic potential.

Taking the potential reference at the bulk, the hole concentration is given by

$$p = n_i e^{u_F - u} . \quad (2.4.9)$$

The hole quasi-Fermi potential reduces to the equilibrium Fermi potential. Deep in the bulk $\phi=0$ and the hole density is at its equilibrium value $p_0 = n_i \exp(u_F)$.

The electron concentration in the channel is given by

$$n = n_i e^{u - u_F - u_C} \quad (2.4.10)$$

where $u_C = V_C / \phi_t$ is the normalized reverse bias of the source-substrate junction. The electron quasi-Fermi potential $\phi_n = \phi_p + V_C = \phi_F + V_C$.

Summarizing, in the three-terminal MOS structure the hole density is controlled by the bulk potential drop ϕ , while the electron density is controlled by the channel-to-source potential $\phi - V_C$.

The resolution of the Poisson-Boltzmann equation for the three-terminal MOS structure is analogous to that of the two-terminal structure. The carrier densities are now given by (2.4.9) and (2.4.10) instead of (2.3.29) and (2.3.30).

The Poisson equation (2.2.1) must be solved with the boundary conditions in (2.2.5). Until recently the standard result, from the work of Pao and Sah [18], was based on writing the bulk charge neutrality condition as

$$N_A - N_D = p_0 - n_0 = n_i (e^{u_F} - e^{-u_F}). \quad (2.4.11)$$

The reason for writing bulk neutrality exactly as for the two-terminal structure is that very deep in the bulk, equilibrium prevails and the quasi-Fermi levels reduce to the Fermi level. Recently, it was perceived that the traditional form of the implicit equation for the surface potential was unphysical and numerically ill-conditioned near the flat-band point at $\phi_s = 0$ [19], [20]. A correct physics-based solution for all values of the surface potential was recently achieved [21], simply by writing the remote bulk neutrality (2.4.11) as

$$N_A - N_D = p_0 - n_0 = n_i (e^{u_F} - e^{-u_F - u_C}). \quad (2.4.12)$$

Equation (2.4.12) is the neutrality condition at the transition between space-charge and neutral regions. Using (2.4.12) instead of (2.4.11) leads to [21]:

$$\frac{d^2\phi}{dx^2} = -\frac{\rho}{\epsilon_s} = -\frac{qn_i}{\epsilon_s} (e^{u_F-u} - e^{u-u_F-u_C} + e^{-u_F-u_C} - e^{u_F}) . \quad (2.4.13)$$

Deep in the bulk $u=0$, and charge neutrality prevails ($\rho=0$), for all values of the source bias u_C .

Integrating (2.4.13) in the same way as (2.2.4) gives

$$\left(\frac{d\phi}{dx} \right)^2 \Big|_{x=0} = \frac{2q\phi_i n_i}{\epsilon_s} \left[e^{u_F} (e^{-u_s} + u_s - 1) + e^{-(u_F+u_C)} (e^{u_s} - u_s - 1) \right] . \quad (2.4.14)$$

Equation (2.4.14) is similar to (2.2.7), the only difference being that the normalized quasi-Fermi potential u_F+u_C of the electrons substitutes for the normalized Fermi potential u_F in the term associated with electrons. Equation (2.4.14) is physically sound, and numerically well-conditioned for all values of the surface potential, including flat-band, for the same reasons as (2.2.7). Using Gauss' law, an expression similar to (2.3.10) for the total charge is achieved as

$$-Q'_C = \text{sign}(u_s) \sqrt{2q\phi_i \epsilon_s n_i} \sqrt{e^{u_F} (e^{-u_s} + u_s - 1) + e^{-(u_F+u_C)} (e^{u_s} - u_s - 1)} . \quad (2.4.15)$$

Using (2.3.34) to relate the surface potential to the applied gate voltage V_G , the surface potential master equation becomes

$$V_G - V_{FB} - \phi_s = \frac{\epsilon_s}{C'_{ox}} F(u_s, u_C) \quad (2.4.16)$$

$$F(u, u_C) = \text{sign}(u) \sqrt{\frac{2q\phi_i n_i}{\epsilon_s} [e^{u_F} (e^{-u} + u - 1) + e^{-(u_F+u_C)} (e^u - u - 1)]} \quad (2.4.17)$$

where $F(u, u_C)$ is the field or Kingston function [22], and $F(u_s, u_C)$ gives the surface electric field.

Finally, the small-signal gate capacitance is

$$C_{gg} = \frac{\partial Q'_G}{\partial V_G} \Big|_{V_C} . \quad (2.4.18)$$

The only difference between the expressions for the capacitances of the three-terminal and two-terminal structures is that in the former, the quasi-Fermi potential of electrons $u_F + u_C$ substitutes the Fermi potential u_F in the terms associated with electrons (compare expressions (2.4.15) and (2.3.10)). The resulting C-V curves with u_C as the parameter are shown in Fig. 2.19.

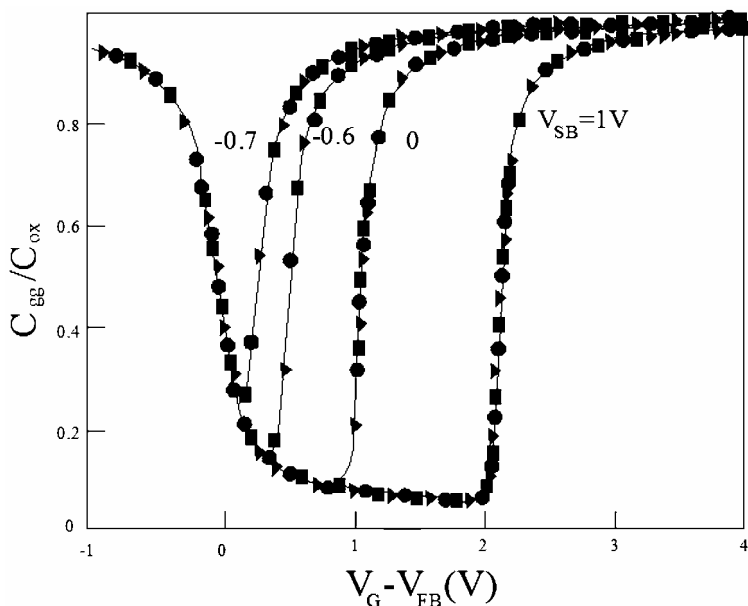


Fig. 2.19 Normalized C-V curves for the three terminal MOS structure. (After [20].)

2.4.3 Small signal equivalent capacitive circuit in inversion

In this section we develop a nearly exact capacitive circuit for the three-terminal structure in inversion (including depletion). In accumulation, the minority carriers are irrelevant, thus the equivalent circuit is the same as that for the two-terminal capacitors.

If $\phi_s > 6\phi_t$ ($\exp(u_s) \gg u_s + 1$ and $\exp(-u_s) \ll u_s - 1$), we can approximate (2.4.15) by

$$-Q'_C = \sqrt{2q\phi_t \epsilon_s n_i e^{u_F}} \sqrt{u_s - 1 + e^{u_s - 2u_F - u_C}} , \quad (2.4.19)$$

and (2.4.16) with (2.4.17) by

$$V_G - V_{FB} - \phi_s = \gamma \sqrt{\phi_s - \phi_t + \phi_t e^{u_s - 2u_F - u_C}} , \quad (2.4.20)$$

where γ is the body effect factor defined in (2.3.23).

For a p-type semiconductor in inversion, the electron and hole capacitances, conventionally called inversion and depletion capacitances, respectively, are defined as

$$C'_i = - \left. \frac{\partial Q'_I}{\partial \phi_s} \right|_{V_C} \quad (2.4.21)$$

$$C'_b = - \left. \frac{\partial Q'_B}{\partial \phi_s} \right|_{V_C} , \quad (2.4.22)$$

where $Q'_I = Q'_N$ and $Q'_B = Q'_P$ are the total stored electron and hole charges, defined in (2.3.12). The expressions for the capacitances in inversion of a three-terminal MOS capacitor are also given by (2.3.14), as in a two-terminal device, and repeated here with the approximate formulas

$$C'_b = C'_p = \frac{q}{F_s} p_0 \quad ; \quad C'_i = C'_n = \frac{q}{F_s} n_s . \quad (2.4.23)$$

In (2.4.23) we have neglected the surface concentration of holes p_s compared to the hole bulk density p_0 and the bulk concentration of electrons n_0 compared to the surface concentration n_s .

Calculating $F_s = -Q'_C / \epsilon_s$ using (2.4.19) we obtain

$$C'_b = \gamma C'_{ox} \frac{1}{2\sqrt{\phi_s - \phi_t + \phi_t e^{u_s - 2u_F - u_C}}} \quad (2.4.24)$$

and

$$C'_i = \gamma C'_{ox} \frac{e^{u_s - 2u_F - u_C}}{2\sqrt{\phi_s - \phi_t + \phi_t e^{u_s - 2u_F - u_C}}} . \quad (2.4.25)$$

Differentiating (2.4.20) with respect to V_C we obtain

$$\left. \frac{\partial \phi_s}{\partial V_C} \right|_{V_G} = \frac{\frac{\gamma e^{u_s - 2u_F - u_C}}{2\sqrt{\phi_s - \phi_t + \phi_t e^{u_s - 2u_F - u_C}}}}{1 + \frac{\gamma [1 + e^{u_s - 2u_F - u_C}]}{2\sqrt{\phi_s - \phi_t + \phi_t e^{u_s - 2u_F - u_C}}}}. \quad (2.4.26)$$

Using the expressions of the depletion and inversion capacitances given in (2.4.24) and (2.4.25), equation (2.4.26) is compactly rewritten in the form:

$$\left. \frac{\partial \phi_s}{\partial V_C} \right|_{V_G} = \frac{C'_i}{C'_i + C'_b + C'_{ox}}. \quad (2.4.27)$$

This result can be easily interpreted using the three-terminal small signal capacitive equivalent circuit of the gate controlled diode shown in Fig. 2.20.

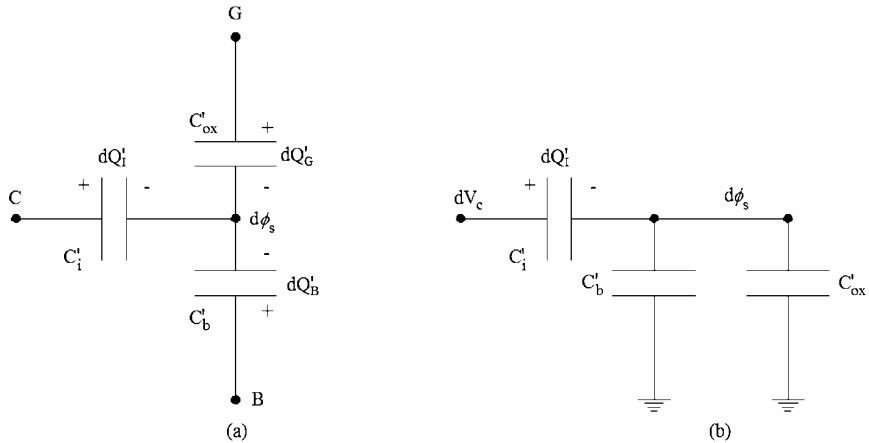


Fig. 2.20 (a) Capacitive model of the three-terminal MOS capacitor, (b) Small-signal equivalent circuit for constant V_G and V_B . The inversion capacitance C'_i is in series with the effective capacitance $C'_b + C'_{ox}$.

To determine the equivalent inversion charge capacitance we can calculate

$$\left. \frac{\partial Q'_I}{\partial V_C} \right|_{V_G} = \left. \frac{\partial Q'_I}{\partial \phi_s} \right|_{V_G} \left. \frac{\partial \phi_s}{\partial V_C} \right|_{V_G}. \quad (2.4.28)$$

From (2.3.34) we can write

$$Q'_I = -C'_{ox}(V_G - V_{FB} - \phi_s) - Q'_B. \quad (2.4.29)$$

Thus,

$$\left. \frac{\partial Q'_I}{\partial \phi_s} \right|_{V_G} = C'_{ox} - \left. \frac{\partial Q'_B}{\partial \phi_s} \right|_{V_G} = C'_{ox} + C'_b. \quad (2.4.30)$$

Combining (2.4.27), (2.4.28) and (2.4.30) we obtain

$$\left. \frac{\partial Q'_I}{\partial V_C} \right|_{V_G} = \frac{(C'_b + C'_{ox})C'_i}{C'_i + C'_b + C'_{ox}}. \quad (2.4.31)$$

As shown in the small signal equivalent circuit of Fig. 2.20, the inversion charge is stored in the capacitor equivalent to the series association of the inversion capacitance with the parallel connection of the oxide and depletion capacitances.

2.4.4 Basic approximations for compact modeling

To obtain compact models of the MOS structures, some approximations are necessary, and the two most important ones are the object of this section.

2.4.4.1 The charge-sheet approximation [23]

The charge-sheet approximation ignores the potential drop across the inversion layer for the calculation of the bulk charge density Q'_B . According to the charge-sheet approximation, Q'_B is given by

$$Q'_B = -\text{sign}(\phi_s) C'_{ox} \gamma \sqrt{\phi_s + \phi_t (e^{-\phi_s/\phi_t} - 1)}. \quad (2.4.32)$$

Equation (2.4.32) can be obtained by integrating the Poisson equation and disregarding the minority carriers or, alternatively, disregarding the electron related terms in (2.4.15). Expression (2.4.32) gives a continuous

model from accumulation through depletion to strong inversion. From (2.3.34) the inversion charge density $Q'_I = Q'_C - Q'_B$ is expressed as

$$Q'_I = -C'_{ox} \left(V_G - V_{FB} - \phi_s + \frac{Q'_B}{C'_{ox}} \right). \quad (2.4.33)$$

For ϕ_s -based models, one can calculate ϕ_s iteratively using the surface potential master equation (2.4.16)-(2.4.17) and the resultant value is used to calculate Q'_B from (2.4.32) and Q'_I from (2.4.33). Q'_G results from charge neutrality $Q'_G = -Q'_B - Q'_I$.

In inversion (including depletion) the exponential term can be neglected and (2.4.32) reduces to

$$Q'_B = -C'_{ox} \gamma \sqrt{\phi_s - \phi_t}. \quad (2.4.34)$$

Using the charge-sheet approximation, a compact formulation of the inversion capacitance in terms of the charges is obtained. Using (2.4.19) and (2.4.34), (2.4.25) can be rewritten as

$$C'_i = -\frac{Q'^2_C - Q'^2_B}{2\phi_t Q'_C}. \quad (2.4.35)$$

Since $Q'_I = Q'_C - Q'_B$, Eq. (2.4.35) is equivalent to

$$C'_i = -\frac{Q'_I}{2\phi_t} \left(1 + \frac{Q'_B}{Q'_B + Q'_I} \right). \quad (2.4.36)$$

The expression for the depletion capacitance in terms of the charge densities results from (2.4.19) and (2.4.24). Writing the total semiconductor charge density $Q'_C = Q'_I + Q'_B$ we obtain

$$C'_b = -\frac{\gamma^2 C'^2_{ox}}{2(Q'_B + Q'_I)}. \quad (2.4.37)$$

Substituting (2.4.36) and (2.4.37) into (2.4.27) we obtain¹

$$\frac{dV_C}{d\phi_s} = 1 - \frac{C'_{ox} \phi_t}{Q'_I} \left[1 + \frac{Q'_I - \gamma^2 C'_{ox}}{Q'_I + 2Q'_B} \right]. \quad (2.4.38)$$

¹ In (2.4.38) we have substituted the partial derivative symbol by the total derivative one for the sake of simplicity. In effect, for constant V_G , ϕ_s is a function of V_C only.

Equation (2.4.38) is the basic relationship used by van de Wiele in [24] to develop the MOS transistor theory. Writing the charges in Eq. (2.4.38) in terms of the surface potential we obtain expression (2.4.39), first presented in [25],

$$\frac{dV_C}{d\phi_s} = 1 + \frac{2(V_G - V_{FB} - \phi_s) + \gamma^2}{(V_G - V_{FB} - \phi_s)^2 - \gamma^2(\phi_s - \phi_t)} \phi_t. \quad (2.4.39)$$

Summarizing, (2.4.27), (2.4.38), and (2.4.39) are equivalent expressions for the derivative $dV_C/d\phi_s$. Clearly, in strong inversion $|Q'_I| \gg C'_{ox}\phi_t$ and $dV_C/d\phi_s \cong 1$, a result that is classically used in the strong inversion MOSFET model. However, in weak and moderate inversion the approximation $dV_C/d\phi_s \cong 1$ no longer holds and the complete expressions for $dV_C/d\phi_s$ are essential to develop a general model of the MOS transistor.

2.4.4.2 The approximate linear relationship between inversion charge density and surface potential

Most recent compact models [26], [27], [28], [29], [30] make use of the linearization of the inversion charge-surface potential relationship since this approximation reduces substantially the analytical intricacies of the MOS equations.

Using (2.4.34) in (2.4.33), the inversion charge becomes

$$Q'_I = -C'_{ox} \left(V_G - V_{FB} - \phi_s - \gamma \sqrt{\phi_s - \phi_t} \right). \quad (2.4.40)$$

Expanding (2.4.40) in power series about ϕ_{sa} (value of surface potential deep in weak inversion, neglecting channel charge) we obtain, for constant V_G [27]:

$$Q'_I \cong C'_{ox} n (\phi_s - \phi_{sa}), \quad (2.4.41)$$

where ²

$$n = 1 + \frac{C'_b|_{\phi_s=\phi_{sa}}}{C'_{ox}} = 1 + \frac{\gamma}{2\sqrt{\phi_{sa} - \phi_t}}. \quad (2.4.42)$$

² We use the symbol n to represent both the electron concentration and the slope factor following the customary notation.

C'_b is the depletion capacitance calculated assuming the inversion charge to be negligible and n is the slope factor, slightly dependent on the gate voltage. The expression for ϕ_{sa} is obtained by disregarding the inversion charge term Q'_l in (2.4.40), resulting in

$$V_G - V_{FB} = \phi_{sa} + \gamma \sqrt{\phi_{sa} - \phi_t} \quad (2.4.43)$$

or

$$\sqrt{\phi_{sa} - \phi_t} = \sqrt{V_G - V_{FB} - \phi_t + \frac{\gamma^2}{4} - \frac{\gamma}{2}}. \quad (2.4.44)$$

Taking the derivative of Eq. (2.4.43) with respect to ϕ_{sa} , it follows that

$$n = \frac{dV_G}{d\phi_{sa}}. \quad (2.4.45)$$

Fig. 2.21 shows the dependence of Q'_l on ϕ_s . At constant V_G , the linearization of Q'_l with respect to ϕ_s , according to Eq. (2.4.41), fits very well the charge-sheet model, as readily seen in Fig. 2.21.

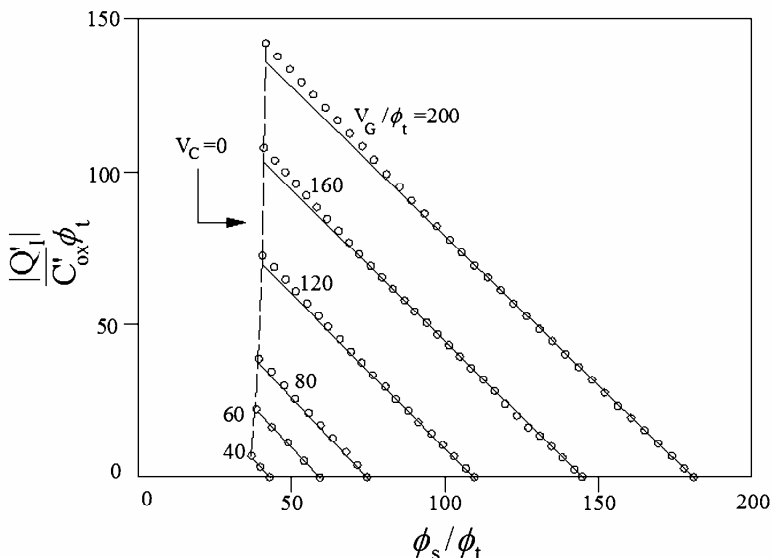


Fig 2.21 Normalized inversion charge density vs. normalized surface potential computed from: (—) the linearized expression (2.4.41), (ooooooo) the theoretical implicit equations (2.4.40) and (2.4.16).

2.4.5 The pinch-off voltage

The channel charge density corresponding to the effective capacitance ($C'_{ox} + C'_b = nC'_{ox}$) times the thermal voltage, or thermal charge [26], will be used as the normalization charge and as a reference to define pinch-off [31]

$$Q'_{IP} = -nC'_{ox}\phi_t. \quad (2.4.46)$$

The name pinch-off is retained herein for historical reasons and means the channel potential corresponding to a small (but well-defined) amount of carriers in the channel.

The channel-to-substrate voltage (V_C) for which the channel charge density equals Q'_{IP} is called the pinch-off voltage V_P . The equilibrium threshold voltage V_{T0} , measured for $V_C=0$, is the gate voltage for which the channel charge density equals Q'_{IP} or, in others words, is the gate voltage for which the pinch-off voltage is zero. Using approximation (2.4.41) to calculate the surface potential ϕ_{sp} at pinch-off gives

$$\phi_{sp} = \phi_{sa} - \phi_t. \quad (2.4.47)$$

In order to calculate the pinch-off voltage, we can substitute (2.4.20) into (2.4.40), yielding

$$Q'_I = -\gamma C'_{ox} \left[\sqrt{\phi_s + \phi_t e^{(\phi_s - 2\phi_F - V_C)/\phi_t}} - \phi_t - \sqrt{\phi_s - \phi_t} \right]. \quad (2.4.48)$$

Substituting ϕ_s , V_C , and Q'_I with ϕ_{sp} , V_P , and Q'_{IP} , respectively, in (2.4.48) and linearizing (2.4.48) around ϕ_{sp} results in

$$V_P = \phi_{sa} - 2\phi_F - \phi_t \left(1 + \ln \left(\frac{n}{n-1} \right) \right). \quad (2.4.49)$$

Fig. 2.22 shows plots of both the pinch-off voltage and the slope factor against the gate voltage. The pinch-off voltage varies almost linearly with the gate voltage while the slope factor changes only slightly for large gate voltage variations. For hand design, it is useful to have a simplified expression for the pinch-off voltage [32]. Neglecting the third term in (2.4.49), of the order of the thermal voltage ϕ_t , the conventional expression of V_P follows as

$$V_P = \phi_{sa} - 2\phi_F. \quad (2.4.50)$$

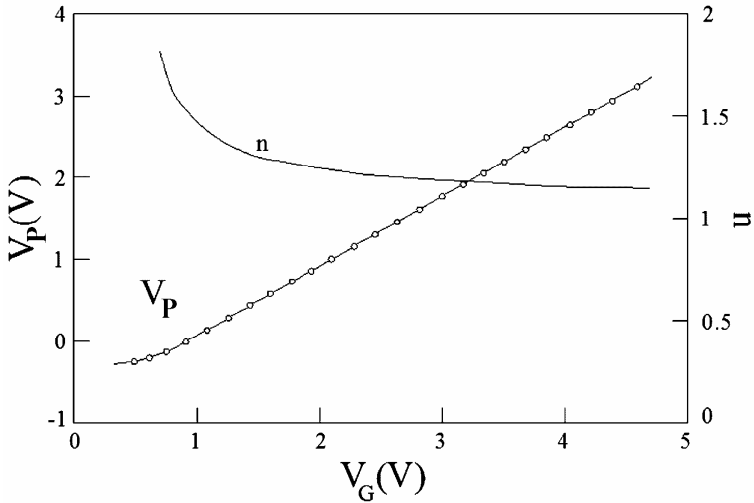


Fig. 2.22 Pinch-off voltage and slope factor vs. gate voltage for an NMOS transistor with $t_{ox}=280 \text{ \AA}$ and $W=L=25 \mu\text{m}$: (o) experimental values of V_P ; (—) values of V_P and n calculated from theoretical expressions (2.4.42) and (2.4.50). (After [33].)

The equilibrium threshold voltage V_{T0} , as defined above, corresponds to the condition $V_P=0$ or

$$\phi_{sa} = 2\phi_F + \phi_t \left(1 + \ln \left(\frac{n}{n-1} \right) \right). \quad (2.4.51)$$

From the definition of ϕ_{sa} in (2.4.43) we obtain

$$\begin{aligned} V_{T0} &= V_{FB} + 2\phi_F + \phi_t \left(1 + \ln \left(\frac{n}{n-1} \right) \right) + \gamma \sqrt{2\phi_F + \phi_t \ln \left(\frac{n}{n-1} \right)} \\ &\cong V_{FB} + 2\phi_F + \gamma \sqrt{2\phi_F}. \end{aligned} \quad (2.4.52)$$

The approximation in (2.4.52) is the classical definition of the threshold voltage. The linear approximation of V_P can be written as

$$V_P \cong \frac{V_G - V_{T0}}{n} \quad (2.4.53)$$

because using (2.4.50) and (2.4.45) the slope dV_P/dV_G is given by

$$\frac{dV_P}{dV_G} = \frac{d\phi_{sa}}{dV_G} = \frac{1}{n}. \quad (2.4.54)$$

For hand calculations, n can be assumed as constant for several decades of current.

2.4.6 The unified charge control model (UCCM)

To develop an explicit expression for the inversion charge in terms of the applied voltages, we will consider the equivalent small signal capacitive model of Fig. 2.20 for a fixed gate voltage value. Expression (2.4.31) is rewritten below for convenience as

$$dQ'_I \left(\frac{1}{C'_{ox} + C'_b} + \frac{1}{C'_i} \right) = dV_C. \quad (2.4.55)$$

We have substituted the partial differential terms with total differentials because, for constant V_G , the inversion charge is a function only of the variable V_C . Substituting C'_i and C'_b with expressions (2.4.36) and (2.4.37), which are given in terms of the charge densities, allows a first order differential equation in Q'_I because the depletion charge Q'_B is known in terms of V_G and Q'_I . This equation is rather cumbersome, but a very useful model is obtained using further approximations [33]:

(i) The depletion capacitance per unit area is assumed to be constant along the channel and is calculated assuming the inversion charge to be negligible in the potential balance equation;

(ii) The inversion capacitance is given by

$$C'_i = -\frac{Q'_I}{\phi_i}. \quad (2.4.56)$$

In weak, moderate and not very strong inversion, the inversion capacitance given by (2.4.36) can be approximated by (2.4.56). In very strong inversion (2.4.56) overestimates the inversion capacitance by a factor of 2, but in this region $C'_i \gg C'_{ox}, C'_b$ and, therefore, its exact value is not relevant to calculate the total equivalent capacitance in

Eq. (2.4.55). Using the hypotheses (i) and (ii) above allows us to obtain, from (2.4.55), the following relationship

$$dQ'_I \left(\frac{1}{nC'_{ox}} - \frac{\phi_t}{Q'_I} \right) = dV_C. \quad (2.4.57)$$

Integrating (2.4.57) from an arbitrary channel potential V_C to the pinch-off potential V_P yields the UCCM as

$$V_P - V_C = \phi_t \left[\frac{Q'_{IP} - Q'_I}{nC'_{ox}\phi_t} + \ln \left(\frac{Q'_I}{Q'_{IP}} \right) \right], \quad (2.4.58)$$

where Q'_{IP} is the value of Q'_I at pinch-off. So far, in the derivation of (2.4.58) we have not defined the value of Q'_{IP} . However, for asymptotic correctness of (2.4.58) in weak inversion, one should use $Q'_{IP} = -nC'_{ox}\phi_t$, which coincides with the value we previously defined for the pinch-off charge.

Even though the UCCM has been presented as a semi-empirical model [34], it has been shown that the UCCM can be readily derived from the Boltzmann–Poisson equation using two approximations [33]. A fundamental property of (2.4.58) is that, in weak inversion, it is asymptotically coincident with the surface potential charge-sheet model. Substituting (2.4.46) and (2.4.49) into (2.4.58) and considering $Q'_I / Q'_{IP} \ll 1$, it follows that

$$Q'_I = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\phi_{sa} - \phi_t}} \phi_t e^{(\phi_{sa} - 2\phi_F)/\phi_t} e^{-V_C/\phi_t}. \quad (2.4.59)$$

Expression (2.4.59) is also the asymptotic expression in weak inversion for the calculation of the inversion charge using equations (2.4.20), (2.4.33) and (2.4.34).

Substituting the first order approximation of V_P of (2.4.53) into (2.4.58) gives

$$Q'_{IP} - Q'_I + nC'_{ox}\phi_t \ln \left(\frac{Q'_I}{Q'_{IP}} \right) = C'_{ox} (V_G - V_{T0} - nV_C). \quad (2.4.60)$$

Equation (2.4.60) is useful for hand analysis, but the exponential dependence of Q'_I on V_P in weak inversion precludes (2.4.60) or

expressions based on approximations of the pinch-off voltage from being used for accurate modeling.

2.4.7 Comparison between charge-sheet model and UCCM

The numerical solution of the Poisson equation, (2.4.13), will be used as a baseline for comparison of charges. Expression (2.4.13) is solved by finite differences with a non-uniform discrete mesh in x . Equations (2.4.16) and (2.4.17) with $u=u_s$ provide the boundary conditions for solving (2.4.13).

In Fig. 2.23, we illustrate the results of the inversion charge density obtained numerically, using the charge-sheet model, and UCCM. UCCM and the ϕ_s -based model give very good approximations for the inversion charge densities from weak, through moderate to strong inversion. For the simulations in Figs. 2.23, 2.24, and 2.25, the following parameters were used (unless specified otherwise): temperature = 27 °C, oxide thickness = 2nm, doping concentration = $2E18 \text{ cm}^{-3}$.

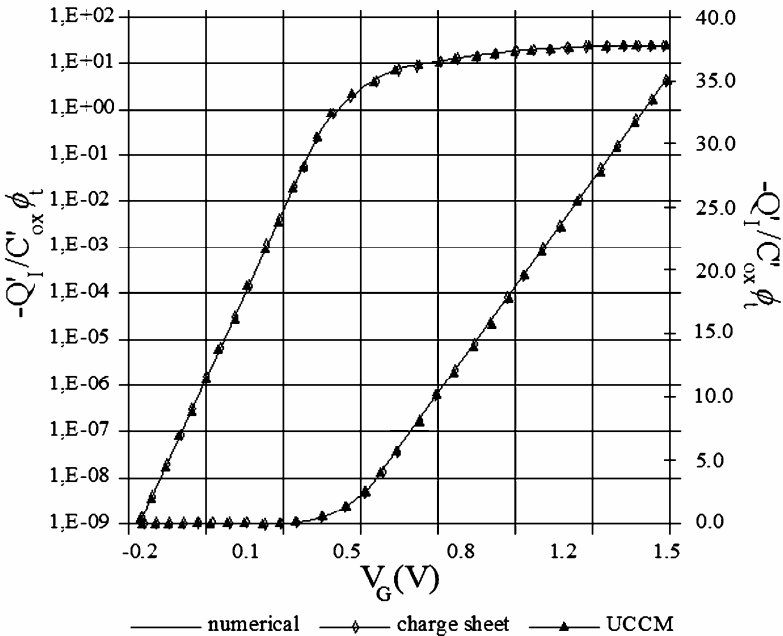


Fig. 2.23 Inversion charge density vs. gate voltage.

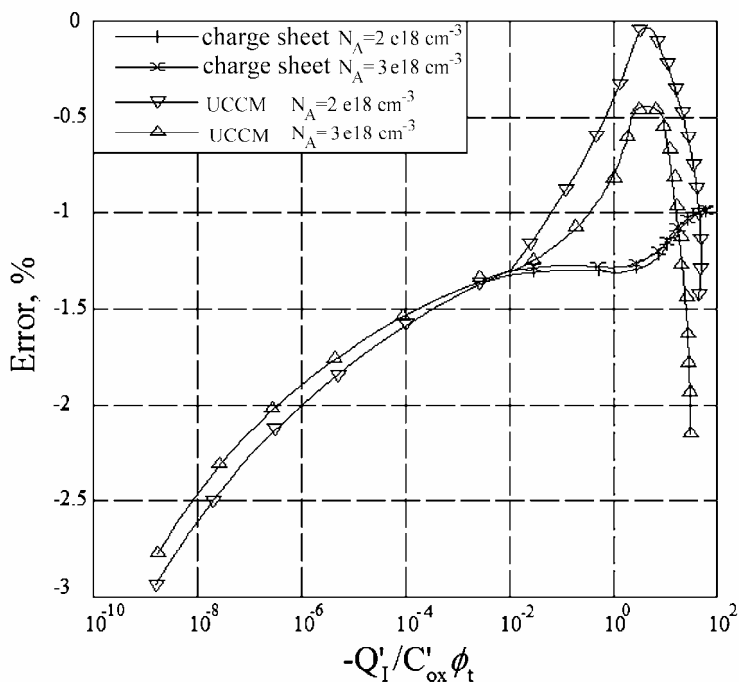


Fig. 2.24 Error in inversion charge density for two doping concentrations. The inversion charge density used as reference is obtained by the numerical resolution of the Poisson equation.

As shown in Figs. 2.24 and 2.25, both ϕ_s -based charge-sheet model and UCCM approaches give errors of less than 3% under normal bias conditions and for a wide range of physical parameters. UCCM and the charge-sheet model are strictly equivalent in weak inversion as expected. An interesting result from Figs. 2.24 and 2.25 is that UCCM gives a better approximation than the charge-sheet model for moderate inversion while in strong inversion the opposite is observed. The better accuracy of the UCCM in moderate inversion is related to the fact that the capacitive model on which the UCCM is based is more general than the charge-sheet approximation.

In strong inversion, the value of the inversion capacitance ($-Q'_I / \phi_t$) used in UCCM is less accurate and the charge-sheet model gives a better result.

All charges are obtained naturally in the ϕ_s -based model, but there is a need for supplementary bulk and/or gate charge models in Q'_I -based approaches.

Summarizing, UCCM and the surface potential calculation give accurate values for the inversion charge in the useful range of MOS transistor operation. Some subtle differences between the inversion charges calculated from UCCM and the charge-sheet model are found.

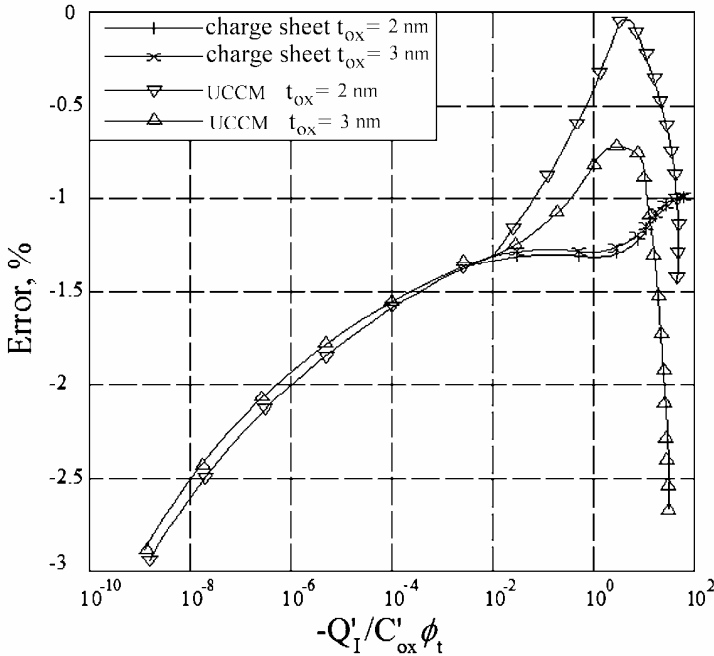


Fig. 2.25 Error in inversion charge density for two oxide thicknesses. The inversion charge density used as reference is obtained by the numerical resolution of the Poisson equation.

2.5 Real C-V curves: Interface traps, polysilicon depletion, and quantum effects

The model of the stored charges in the MOS structure developed in the previous section is somewhat idealized. Thus, the real C-V curves are not exactly as we have presented them. Nevertheless, with slight

modifications to the equivalent capacitive circuit of the MOS structure it is possible to obtain a fair fit of real C-V curves. Three main nonidealities may be included: at the interface, in the gate material, and in the semiconductor.

2.5.1 Interface-trap capacitance

Traps, previously mentioned in Section 2.2, refer to energy levels generated by imperfections within the semiconductor crystal [14], [35], allowed for the carriers in the forbidden energy gap of the semiconductor. These energy levels can change their charge state by exchange of carriers with the semiconductor [8], [14], [35]. Clearly, an interface is a huge defect in a periodic crystal, and interface traps are always present. These traps interact with the conduction band by capturing or emitting electrons and with the valence band by capturing or emitting holes [14]. To derive a compact model of the MOS capacitance that accounts for interface traps, it is usually considered that the number of interface traps N'_{ss} ($\text{V}^{-1}\text{-cm}^{-2}$) per unit potential³ per unit area is constant. As shown in annex G, in inversion, the occupancy of the interface traps is determined by the position of the quasi-Fermi level of the minority carriers. The total charge density stored at the interface is given [36] by

$$Q'_o - qN'_{ss}(\phi_s - V_C). \quad (2.5.1)$$

Q'_o is the effective (converted to the interface) fixed oxide charge density together with the charge stored in the interface traps at flat-band. The second term in (2.5.1) corresponds to the bias dependent trap charge density. The interface trap capacitance per unit area C'_{ss} is simply given by

$$C'_{ss} = qN'_{ss}. \quad (2.5.2)$$

High densities of interface traps can completely distort the C-V curve as seen in Fig. 2.26. The interface state density must be reduced to the order of $10^{10} \text{ V}^{-1}\text{-cm}^{-2}$ or less to obtain predictable and useful MOS

³ In the technical literature, the density of interface traps N'_{ss} is expressed in either $\text{eV}^{-1}\text{-cm}^{-2}$ or $\text{V}^{-1}\text{-cm}^{-2}$. We decided to use $\text{V}^{-1}\text{-cm}^{-2}$ as the unit of N'_{ss} .

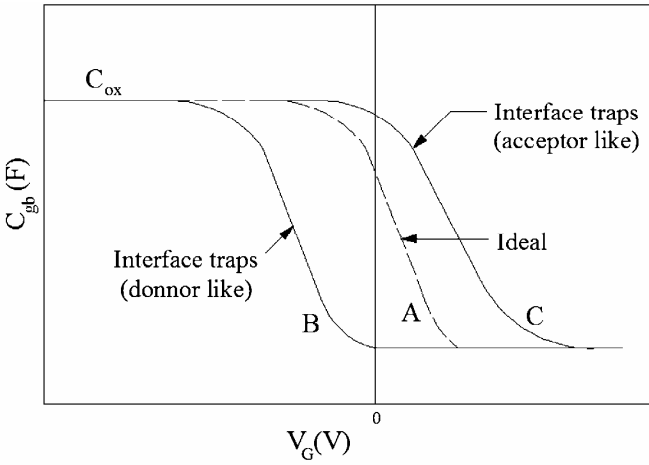


Fig. 2.26 Influence of the interface traps on the high-frequency MOS C-V curve. Curve A is the ideal C-V curve with no traps, curve B is with donnor-like traps, and curve C is with only acceptor-like interface traps. (After [42].)

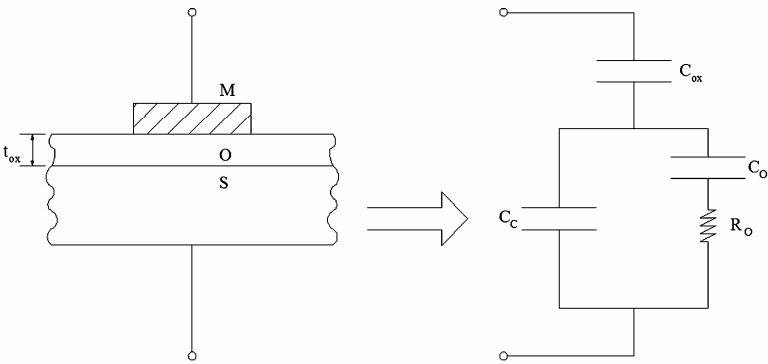


Fig. 2.27 Equivalent circuit of a MOS capacitance considering the equivalent impedance (R_o C_o) of the interface traps. (After [42].)

structures. An accurate model of the interface traps must include a resistive term to model the response time of the traps [14] as seen in Fig. 2.27.

2.5.2 Polysilicon depletion [12]

Polysilicon gates have been used in MOS technologies since the seventies in order to obtain source and drain regions self-aligned to the gate. If the polysilicon doping is not high enough, it cannot be considered equipotential, as is an ideal metal gate.

Processes in which the gates are doped by ion implantation, as is the case of the dual n^+ - p^+ polysilicon gate, are the most affected by the polysilicon depletion [12]. Two structures are the most common, the p-type substrate with n^+ gate and the n-type substrate with p^+ gate. We will analyze the former case.

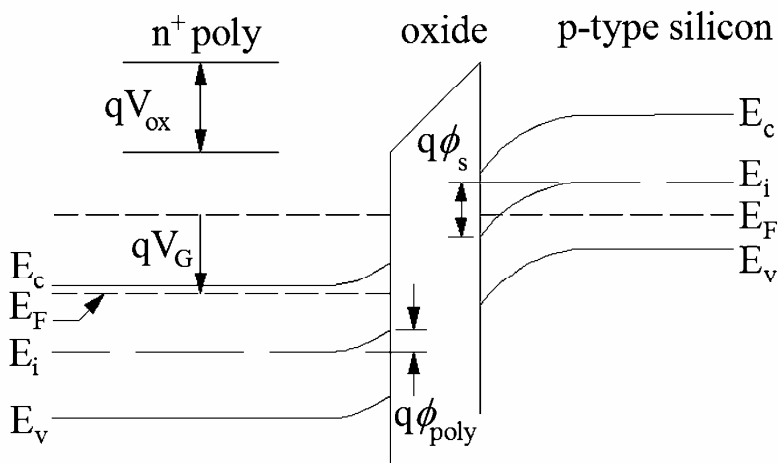


Fig. 2.28 Band diagram showing polysilicon-gate depletion effects when a positive voltage is applied to the n^+ polysilicon gate of a p-type MOS capacitor. (After [12].)

As seen in Fig. 2.28, when the p-semiconductor substrate is in depletion and/or inversion, the positive charge in the polygate is stored in a thin depletion layer. Thus, the capacitance of this depletion layer must be included in the MOS equivalent circuit as will be shown next. Considering the potential drop ϕ_{poly} in the polygate, the potential balance equation (2.3.34) must be rewritten as

$$V_G - V_{FB} = \phi_s + \phi_{poly} - \frac{Q'_C}{C'_{ox}}. \quad (2.5.3)$$

Differentiating (2.5.3) with respect to Q'_C it follows that

$$\frac{1}{C'_{gb}} = \frac{1}{C'_{ox}} + \frac{1}{C'_c} + \frac{1}{C'_{poly}}, \quad (2.5.4)$$

where

$$C'_{poly} = -\frac{dQ'_C}{d\phi_{poly}} = \frac{dQ'_G}{d\phi_{poly}} \quad (2.5.5)$$

is the capacitance of the polysilicon depletion region.

A very simple charge-control model of the capacitances is available when the semiconductor bulk is in inversion. The poly depletion capacitance can be calculated as in (2.3.22) and we obtain

$$C'_{poly} = \frac{\sqrt{2q\epsilon_s N_{poly}}}{2\sqrt{\phi_{poly}}} = \frac{q\epsilon_s N_{poly}}{Q'_G}, \quad (2.5.6)$$

where N_{poly} is the donor doping concentration in the polygate. In strong inversion, a very simple approximate model of the semiconductor capacitance in terms of the stored charge [11] holds,

$$C'_c = -\frac{Q'_C}{2\phi_t} = \frac{Q'_G}{2\phi_t}. \quad (2.5.7)$$

Equation (2.5.7), valid in inversion, is similar to (2.3.19), valid in accumulation. Combining (2.5.4), (2.5.6), and (2.5.7) we obtain

$$\frac{1}{C'_{gb}} = \frac{1}{C'_{ox}} + \frac{2\phi_t}{Q'_G} + \frac{Q'_G}{q\epsilon_s N_{poly}}. \quad (2.5.8)$$

Since the second term decreases with Q'_G and the third increases with Q'_G , a maximum is reached when the two terms are equal, *i.e.*, for

$$Q'_G = \sqrt{2\phi_t q\epsilon_s N_{poly}}. \quad (2.5.9)$$

This local maximum of the capacitance characteristic in inversion is clearly shown in Fig. 2.29. For even higher values of gate voltage, an inversion p-channel in the n^+ polygate may appear and, due to the proximity of the inversion holes to the oxide, the total gate capacitance approaches the oxide capacitance. In Fig. 2.29, in addition to the capacitance reduction, a slight shift in the flat-band voltage is apparent. Considering the semiconductor bulk in depletion or weak inversion, then

$$\begin{aligned} Q'_B &= -C'_{ox} \gamma \sqrt{\phi_s} \\ Q'_G &= C'_{ox} \gamma_{poly} \sqrt{\phi_{poly}}. \end{aligned} \quad (2.5.10)$$

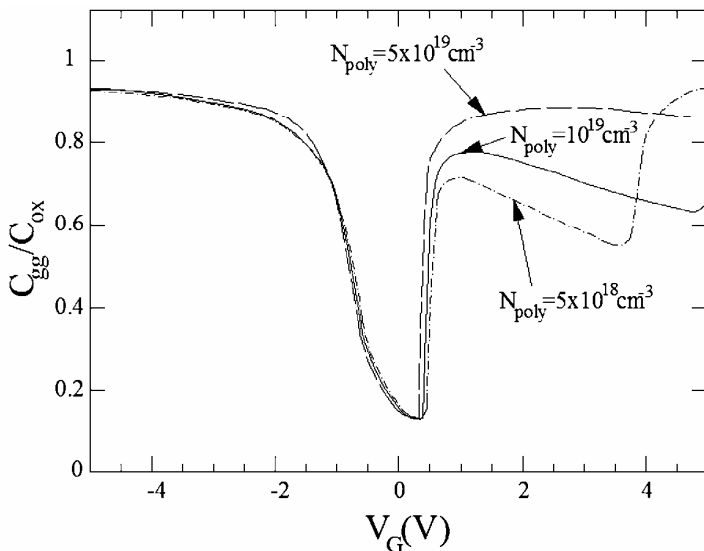


Fig.2.29 Low-frequency C-V curve for a p-type MOS capacitor with n^+ polysilicon gate doped at several different concentrations. (After [37].)

Neglecting the fixed oxide charge we have $Q'_G + Q'_B = 0$. Thus, it follows from (2.5.10) that

$$\frac{\phi_{poly}}{\phi_s} = \left(\frac{\gamma}{\gamma_{poly}} \right)^2 = \frac{N_A}{N_{poly}}, \quad (2.5.11)$$

Writing the inversion charge density from (2.5.3) as

$$Q'_I = -C'_{ox} \left(V_G - V_{FB} - \phi_s - \phi_{poly} - \gamma \sqrt{\phi_s} \right) \quad (2.5.12)$$

and using (2.5.11) and (2.5.12), the increase in both the slope factor and the threshold voltage are readily obtained as

$$n_{poly} = 1 + \frac{1}{2\gamma\sqrt{\phi_{sa}}} + \frac{N_A}{N_{poly}} \quad (2.5.13)$$

$$\Delta V_{T0} = \frac{N_A}{N_{poly}} (2\phi_F).$$

Summarizing, the effects of the gate polysilicon depletion are gate capacitance reduction, slope factor increase and threshold (or flat-band) voltage increase.

2.5.3 Quantum mechanical (QM) effects on the semiconductor capacitance

The scaling down of MOS transistors is accompanied by both thinner oxides and more heavily doped channels. The resulting increase in the transverse electric field at the oxide-semiconductor interface gives rise to quantum confinement in the channel, *i.e.*, carriers in the inversion layer are confined in a narrow potential well, where quantum effects have to be accounted for [38], [39]. In strong inversion, the electric field at the interface can be so strong, particularly for heavily doped substrates, that the carriers are confined to a very short distance from the interface. The potential well is so narrow that the motion of the carriers of the surface channel is quantized in the direction perpendicular to the interface, as shown in Fig. 2.30. This situation is analogous to the classical textbook example of electrons in an infinitely deep potential well. Therefore, the operation of deeply scaled MOS transistors cannot be accurately described by pure classical physics [38], [40]; accurate calculation of the inversion charge requires introducing concepts derived from quantum mechanics (QM).

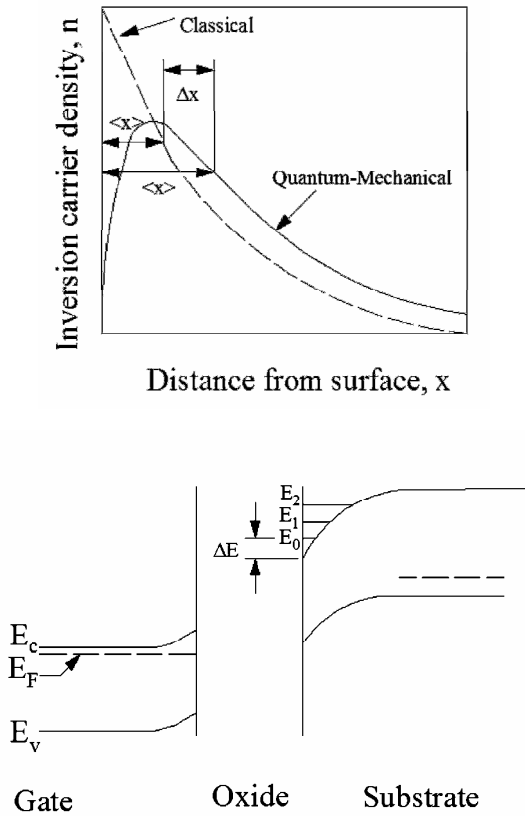


Fig. 2.30 (a) Electron density $n(x)$ as a function of transversal position x for the classical (dashed line) and the quantum-mechanical (solid line) case. (b) Energy-band diagram (in transversal direction) of an n -type MOS transistor in strong inversion. (After [40].)

The energy of a particle in the potential well can have only discrete (quantized) values, called energy levels, and there is a minimum (ground) energy level. For the inversion channel, each quantized energy level of the electron movement in the direction normal to the surface corresponds to a conduction sub-band, because the carriers are free to move parallel to the interface.

In the case of the deep rectangular well, the wave functions representing the electrons are zero at the boundary of the well; consequently the carrier (probability) density is at a maximum inside the

well and not at the boundaries. For the inversion channel the situation is analogous, the peak density is not at the interface but at some distance from it, as illustrated in Fig. 2.30.

Clearly, the quantum effect in field effect devices requires a much more elaborate analytical treatment than that of the rectangular well. One reason is that the depth of the potential well is not infinite and consequently the wave function is not exactly zero outside the well. Also, the depth of the well depends on the strength of the electric field, which decreases going inward in the semiconductor.

The exact analysis of QM effects requires the coupled resolution of the Schrödinger and Poisson equations and does not allow a compact model. Considering a constant electric field, and consequently decoupling the Poisson and Schrödinger equations, greatly simplifies the problem, but even so the calculation of the carrier density is too cumbersome for a compact model. In weak inversion, the inclusion of the various sub-bands in the calculation of the inversion carrier density is not necessary, because the splitting of the energy levels is small compared with the thermal voltage kT/q . In strong inversion, two consequences of quantum effects can be taken into account as semiempirical modifications of the Poisson-Boltzmann equation. First, there is an increase in the value of the surface potential needed to obtain a given electron density because the energy of the lowest sub-band is above that of the bottom of the conduction band. Second, there is a reduction in the value of the inversion capacitance because the peak of the inversion layer is no longer at the semiconductor interface but at some distance from it.

It has been shown [40] that a surface potential ϕ_s^{QM} greater than the conventional value of ϕ_s is required for a given inversion carrier density when QM effects are accounted for, according to

$$\phi_s^{QM} = \phi_s + \Delta E / q + F_s \Delta x, \quad (2.5.14)$$

where ΔE is the energy gap between the bottom of the conduction band and the first allowed energy level, F_s is the electric field perpendicular to the interface, and Δx is the increase in the average distance to the interface compared to the classical solution. Equation (2.5.14) includes

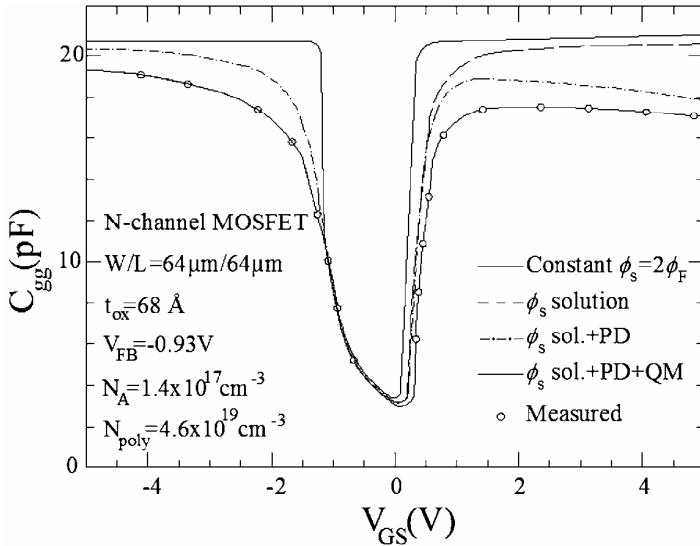


Fig. 2.31 Measured and calculated C-V curves including various physical effects. (After [41].). The curve for constant $\phi_s = 2\phi_F$ represents the classical gate capacitance model where $\phi_s = 0$ in accumulation and $\phi_s = 2\phi_F$ for $V_{GS} > V_T$

both effects of band gap widening and the displacement of the electron distribution. The exact calculation of Δx is not easy, but it has been found using a variational method that $qF_s \Delta x \cong (4/9)\Delta E$. As a result quantum effects are modeled by an effective bandgap increase given by

$$E_g^{QM} = E_g + \frac{13}{9} \Delta E. \quad (2.5.15)$$

Finally, QM effects are accounted for by using a correction for the intrinsic carrier density. The intrinsic carrier concentration in the inversion layer is modified to give

$$n_i^{QM} = n_i e^{-\Delta E_g / 2kT}$$

$$\Delta E_g = \frac{13}{9} \beta \left(\frac{\epsilon_s}{4kT} \right)^{1/3} F_s^{2/3} \quad (2.5.16)$$

$$\beta = 4.1 \times 10^{-8} \text{ eV.cm.}$$

The dependence of the band gap widening on the power 2/3 of the field has been deduced assuming a triangular-shaped potential well. The

value of the constant β has been determined from the voltage shift observed at high doping levels.

The results of the above model are depicted in Fig. 2.31 where the excellent fitting of the experimental data is apparent [41]. Fig. 2.32 compares the surface potential calculated with and without QM corrections [41]. The classical value of $2\phi_F$ for the surface potential in strong inversion is given as a reference. Clearly, QM correction is irrelevant in weak inversion, but in strong inversion the reduction of the semiconductor capacitance, compared to its value in the classical model, has as a consequence an increase in the surface potential.

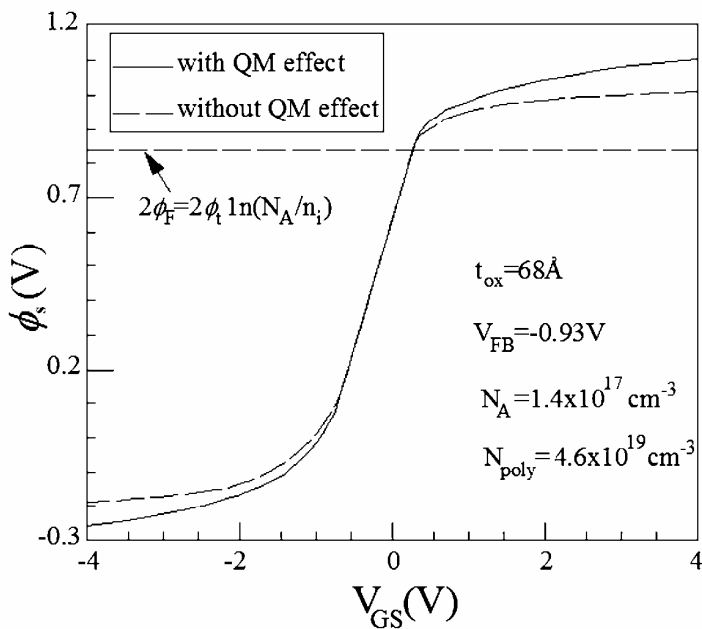


Fig. 2.32 Surface potential solution corresponding to Fig. 2.31. Solid line, with QM effect; dashed line, without QM effect; dotted line, constant $2\phi_F$ level. (After [41].)

Bibliography

- [1] Richard B. Adler, Arthur C. Smith, and Richard L. Longini, *Introduction to Semiconductor Physics*, SEEC Series, vol. 1, John Wiley & Sons, New York, 1964.
- [2] Chih-Tang Sah, *Fundamentals of Solid-State Electronics*, World Scientific, Singapore, 1991.

- [3] W. L. Brown, "n-type surface conductivity on p-type Ge," *Phys. Rev.*, vol. 91, no. 3, pp. 518-527, 1953.
- [4] C. G. B. Garrett and W. H. Brattain, "Physical theory of semiconductor surfaces," *Phys. Rev.*, vol. 99, no. 2, pp. 376-387, 1955.
- [5] Richard S. C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New-York, 1970.
- [6] G. C. Dousmanis and R. C. Duncan, Jr., "Calculation on the shape and extent of space charge regions in semiconductor surfaces," *J. Appl. Phys.*, vol. 29, no. 12, pp. 1627-1629, Dec. 1958.
- [7] S. M. Sze, *Physics of Semiconductor Devices*, Wiley-Interscience, New York, 1969.
- [8] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, New York, 1967.
- [9] Dawon Kahng, "A historical perspective on the development of MOS transistors and related devices," *IEEE Trans. Electron Devices*, vol. 23, no. 7, pp. 655-657, July 1976.
- [10] Robert F. Pierret, *Field Effect Devices*, 2nd ed., Addison-Wesley, Reading, Mass., 1990.
- [11] Alexandre Ternes Behr, Márcio Cherem Schneider, Sidnei Noceti Filho, and Carlos Galup Montoro, "Harmonic distortion caused by capacitors implemented with MOSFET gates," *IEEE J. Solid-State Circuits*, vol. 27, no. 10, pp. 1470-1475, Oct. 1992.
- [12] Yuan Taur and Tak H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.
- [13] E. H. Snow, A. S. Grove, B. E. Deal, and C. T. Sah, "Ion transport phenomena in insulating films," *J. Appl. Phys.*, vol. 36, no. 5, pp. 1664-1673, May 1965.
- [14] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*, Wiley-Interscience, New York, 1982.
- [15] A. S. Grove, E. H. Snow, B. E. Deal, and C. T. Sah, "Simple physical model for the space-charge capacitance of metal-oxide-semiconductor structures," *J. Appl. Phys.*, vol. 35, no. 8, pp. 2458-2460, Aug. 1964.
- [16] Yannis Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed., McGraw-Hill, Boston, 1999.
- [17] William Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Co., Princeton, NJ, 1950.
- [18] H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electron.*, vol. 9, no. 10, pp. 927-937, Oct. 1966.
- [19] Colin C. McAndrew and James J. Victory, "Accuracy of approximations in MOSFET charge models," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 72-81, Jan. 2002.
- [20] Weimin Wu, Ten-Lon Chen, Gennady Gildenblat, and Colin C. McAndrew, "Physics-based mathematical conditioning of the MOSFET surface potential equation," *IEEE Trans. Electron Devices*, vol. 51, no. 7, pp. 1196-1200, July 2004.
- [21] Chih-Tang Sah, "A history of MOS transistor compact modeling", *Proc. Workshop on Compact Modeling, Nanotech 2005*, pp. 347-390.

- [22] Robert H. Kingston and Siegfried F. Neustadter, "Calculation of the space charge, electric field, and free carrier concentration at the surface of a semiconductor," *J. Appl. Phys.*, vol. 26, no. 6, pp. 718-720, June 1955.
- [23] J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electron.*, vol. 21, no. 2, pp. 345-355, Feb. 1978.
- [24] F. van de Wiele, "A long-channel MOSFET model," *Solid-State Electron.*, vol. 22, no. 12, pp. 991-997, Dec. 1979.
- [25] G. Baccarani, M. Rudan, and G. Spadini, "Analytical i.g.f.e.t. model including drift and diffusion currents," *IEEE J. Solid-St. and Electron Dev.*, vol. 2, no. 2, pp. 62-69, Mar. 1978.
- [26] Mary Ann Maher and Carver A. Mead, "A physical charge-controlled model for MOS transistors," in *Advanced Research in VLSI*, P. Losleben (ed.), MIT Press, Cambridge, MA, 1987.
- [27] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electron.*, vol. 38, no. 11, pp. 1945-1952, Nov. 1995.
- [28] Jean-Michel Sallese, Matthias Bucher, François Krummenacher, and Pierre Fazan, "Inversion charge linearization in MOSFET modeling and rigorous derivation of the EKV compact model," *Solid-State Electron.*, vol. 47, no. 4, pp. 677-683, Apr. 2003.
- [29] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "Physical Background of MOS Model 11," Nat. Lab. Unclassified Report 2003/00239, 2003. [Online]. Available: http://www.semiconductors.philips.com/Philips_Models/.
- [30] Gennady Gildenblat, Hailing Wang, Ten-Lon Chen, Xin Gu, and Xiaowen Cai, "SP: An advanced surface-potential-based compact MOSFET model," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1394-1406, Sep. 2004.
- [31] Carlos Galup-Montoro, Márcio Cherem Schneider, and Ana Isabela Araújo Cunha, "A current-based MOSFET model for integrated circuit design," Chapter 2 of *Low-Voltage/Low-Power Integrated Circuits and Systems*, pp. 7-55, edited by Edgar Sánchez-Sinencio and Andreas G. Andreou, IEEE Press, Piscataway, NJ, 1999.
- [32] Christian C. Enz, François Krummenacher, and Eric A. Vittoz "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *J. Analog Integr. Circuits and Signal Process.*, vol. 8, pp. 83-114, July 1995.
- [33] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "Derivation of the unified charge control model and parameter extraction procedure," *Solid-State Electron.*, vol. 43, no. 3, pp. 481-485, Mar. 1999.
- [34] Young Hee Byun, Kwyro Lee, and Michael Shur, "Unified charge control model and subthreshold current in heterostructure field effect transistors," *IEEE Electron Device Lett.*, vol. 11, no. 1, pp. 50-53, Jan. 1990.
- [35] W. Shockley and W. T. Read, Jr., "Statistics of recombination of holes and electrons," *Phys. Rev.*, vol. 87, no. 5, pp. 835-842, Sep. 1952.
- [36] Roger J. Van Overstraeten, Gilbert J. Declerck, and Paul A. Muls, "Theory of the MOS transistor in weak inversion - New method to determine the number of surface states," *IEEE Trans. Electron Devices*, vol. 22, no. 5, pp. 282-288, May 1975.

- [37] Rafael Rios and Narain D. Arora, "Determination of ultra-thin gate oxide thicknesses for CMOS structures using quantum effects," in *IEDM Tech. Dig.*, 1994, pp. 613-316.
- [38] Frank Stern, "Quantum properties of surface space-charge layers," *CRC Crit. Rev. Solid State Sci.*, pp. 499-514, May 1974.
- [39] G. Ghibaudo, and R. Clerc, "Characterization and modeling issues in MOS structures with ultra thin oxides," *Proc. 24th Int. Conf. Microelectronics (MIEL 2004)*, pp. 103-114.
- [40] M. J. van Dort, P. H. Woerlee, and A. J. Walker, "A simple model for quantization effects in heavily-doped silicon MOSFETs at inversion conditions," *Solid-State Electron.*, vol. 37, no. 3, pp. 411-414, Mar. 1994.
- [41] Rafael Rios, Narain D. Arora, Cheng-Liang Huang, Nadim Khalil, John Faricelli, and Len Gruber, "A physical compact MOSFET model, including quantum mechanical effects for statistical circuit design applications," *IEDM Tech. Dig.*, 1995, pp. 937-940.
- [42] Narain Arora, "*MOSFET Models for VLSI Circuit Simulation Theory and Practice*," Springer-Verlag, Wien, 1993.

Problems

2.1. For a high-low P⁺P junction in equilibrium, $N_A = N_A^+$ for $x < 0$ and $N_A = N_A^-$ for $x \geq 0$. (a) Prove that the internal potential barrier is $\phi_b = \phi(x = \infty) - \phi(x = -\infty) = \phi_t \ln(N_A^+/N_A^-)$ and the potential drop ϕ_0 in the N_A^+ region is $\frac{\phi_0}{\phi_t} = 1 - \frac{\phi_b/\phi_t}{e^{\phi_b/\phi_t} - 1}$. Plot the curve $\phi_0(\phi_b)$ and determine the limits of variation of ϕ_0 . (b) For $N_A^+ = 10^{17} \text{ cm}^{-3}$ and $N_A^- = 10^{15} \text{ cm}^{-3}$, plot the majority carrier profile.

2.2. For an MOS capacitor, assume $p_0 = N_A$ and $n_0 = n_i^2/N_A$. (a) Show that $Q'_C = \mp \sqrt{2q\epsilon_s N_A} \sqrt{\phi_t \left[(e^{-u_s} + u_s - 1) + e^{-2u_F} (e^{u_s} - u_s - 1) \right]}$; (b) Determine the analytical expression of the low-frequency semiconductor capacitance $C'_c = -dQ'_C/d\phi_s$; (c) Plot $C'_{gb}(V_{GB})$ assuming $V_{FB} = 0$, $t_{ox} = 100 \text{ \AA}$, $\epsilon_{ox}/\epsilon_o = 4$, $\epsilon_s/\epsilon_o = 12$ (ϵ_o is the permittivity of free space), for three different values of substrate doping,

$N_A = 10^{16}, 10^{17}, 10^{18} \text{ cm}^{-3}$. (Hint: $V_{GB} - V_{FB} = \phi_s - Q'_C / C'_{ox}$; $1/C'_{gb} = 1/C'_c + 1/C'_{ox}$).

2.3. Consider the UCCM given by

$$V_P - V_C = \phi_t \left[\frac{Q'_I - Q'_{IP}}{Q'_{IP}} + \ln \left(\frac{Q'_I}{Q'_{IP}} \right) \right]. \quad (\text{P2.3.1})$$

“Regional” models for weak (WI) inversion or strong (SI) inversion can be derived from UCCM by dropping either the linear term or the \ln term in the charge density, respectively. (a) Calculate the value of the inversion charge density, normalized to $Q'_{IP} = -nC'_{ox}\phi_t$, for which the value of the charge density calculated from the WI approximation differs from that calculated from UCCM by 25%; (b) Determine the value of the inversion charge density, normalized to Q'_{IP} , for which the value of the charge density calculated from the SI approximation differs from that calculated from UCCM by 25%. (c) Sketch UCCM as well as both the WI and SI approximations in: (c1) The $\log(Q'_I/Q'_{IP})$ vs V_G plot at constant V_C ; (c2) The $\log(Q'_I/Q'_{IP})$ vs V_C plot at constant V_G . Assume that $V_P \cong (V_G - V_{T0})/n$, with n being constant. (d) Determine the expression for the charge density in terms of both the gate voltage and the channel voltage for: (d1) The WI approximation, (d2) The SI approximation. (e) Calculate the sensitivity of the inversion charge density wrt the gate voltage and wrt the channel voltage. What are the values of the sensitivities for the asymptotic cases of WI and SI? (f) If $V_C = 0$ and $n=1.25$ calculate the normalized inversion charge density for $V_G = V_{T0} = 0.5 \text{ V}$ and $V_G = 0 \text{ V}$. (g) Recalculate the normalized charge densities for the same data in (f) but with $n=1$.

This page intentionally left blank

Chapter 3

The Long-Channel MOSFET: Theory and dc Equations

The derivation of the direct current (dc) equations of a device is one of the main objectives in compact modeling. For long-channel MOS transistors a complete theory is available, which is the subject of this chapter. Recent compact models are based on this theory, but until now the most popular compact models have been based on a threshold voltage (V_T) formulation. This class of models relies on approximate solutions that are only valid in particular regions of operation connected mathematically to provide continuous solutions. There is, therefore, inaccuracy between regions, which generates inaccuracy in simulation results where transition between regions is important to the circuit operation. For these reasons V_T -based models are being substituted by surface potential and charge-based models. This chapter provides an overview of the basic physics that must be modeled to build a compact dc model for the metal-oxide field effect transistor (MOSFET) and describes the fundamentals of several advanced models. In the final sections of the chapter, regional (V_T) models are derived as particular cases to highlight the link between next and past generation compact models. Furthermore, V_T -based models are useful for developing insight into the operation of the transistor and for back-of-the-envelope circuit analysis.

3.1 A brief history of MOSFET theory

Electrical characterization and modeling of the dc current characteristics (Fig. 3.1) began just after the invention of the modern silicon MOS

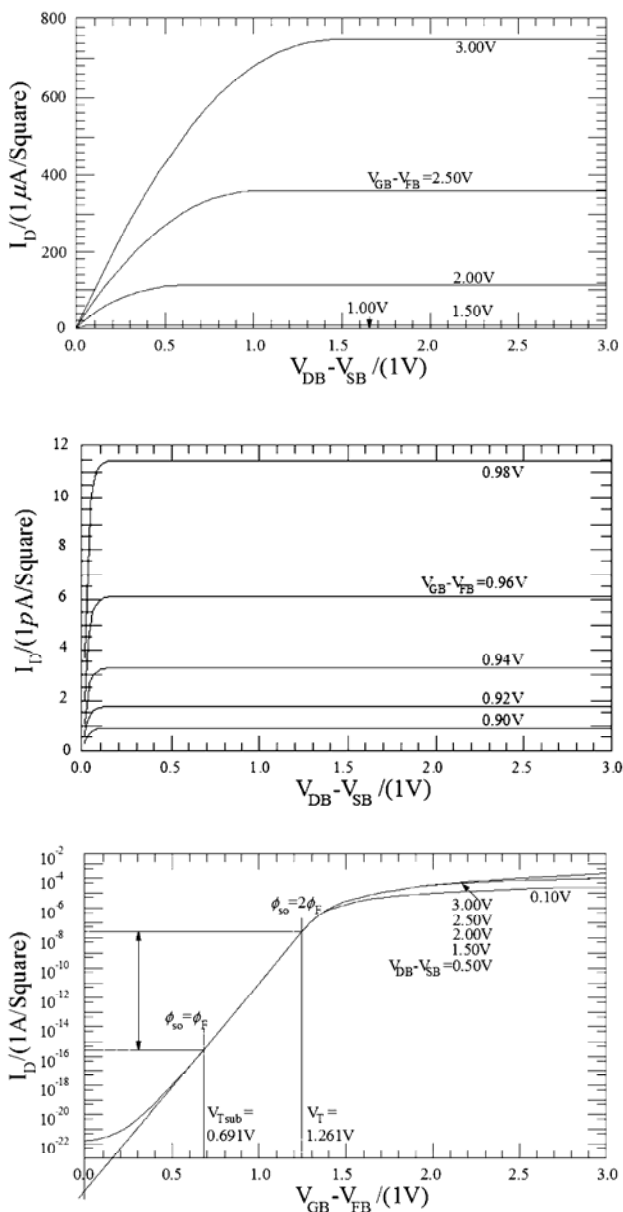


Fig. 3.1 From top to bottom: MOSFET strong inversion output characteristics, weak inversion output characteristics and transfer characteristic. Transistor data: $n_i = 1.33 \times 10^{10} \text{ cm}^{-3}$, $\mu_n = 400 \text{ cm}^2/\text{V}\cdot\text{s}$, $N_A = 10^{18} \text{ cm}^{-3}$, $t_{ox} = 2.0 \text{ nm}$, $V_{SB} = 0$. (After [5].)

transistor in 1960 [1], [2]. The now classical above threshold or strong inversion (SI) theory of the MOSFET was developed in two journal papers in 1964 [3], [4].

In the SI model only the drift component of the drain current is considered. During the 1960s the strong inversion model was refined, and compact models for computer circuit simulation were developed [6]. The complete SI theory was summarized in a 1970 book [7]. In the 1970s, because of the development of low power circuits, the below threshold or weak inversion (WI) mode of operation was explored. The first papers addressing the subthreshold exponential regime appeared in 1972 [8], [9], [10]. Weak inversion models were improved in the 1970s [11], [12]. WI models consider only the diffusion component of the drain current. The first MOSFET theory valid in all operating regions was the Pao-Sah drift-diffusion double-integral model [13]. The Pao-Sah model gives the drain current as a double integral (over the thickness of the inversion layer and the length of the channel). Because of the need for the numerical double integration, it is not considered a compact model. However, the so-called Pao-Sah equation is highly physical and still serves as a “golden” reference to test the accuracy of compact MOSFET models [14], [15]. The first truly compact one-equation all-region models appeared in 1978 [16], [17]. In these two papers, the surface potential is used as the key variable, that is, all the MOSFET characteristics are expressed as functions of the values of the surface potential at the source and drain ends. Brews’ [17] charge-sheet approximation, which assumes that the inversion layer is infinitesimally thin, was appealing, even if his model derivation and the relationship with the more fundamental Pao-Sah model was not clear. Baccarani’s work [16] was rigorously developed, but because of the complexity of the final expressions, Brews’ approach was the preferred one, and he is usually considered to be the originator of the charge-sheet models. In 1979, the links between the Pao-Sah, Brews and Baccarani models were clearly established, with the presentation of a rigorous analytical derivation of the so-called charge-sheet models [18].

Given that the surface potential needs to be calculated numerically with high precision, charge-sheet surface potential models were not considered at that time appropriate for compact modeling. For SPICE

models, the threshold voltage approach [19], [20] was adopted from the very beginning and continued to be until recently. In the threshold voltage approach separate solutions are available for different regions of MOSFET operation (Fig. 3.2). Early SPICE MOSFET models were piecewise formulations, with separate equations used to model different regions of operation. Later models used mathematical techniques to make the models single-piece and a single set of model equations was applicable to all regions of operation [36]. Since V_T -based models use mathematical smoothing functions to describe the transition between weak and strong inversion, they are not accurate enough to represent the moderate inversion region, widely employed in low supply voltage circuits. Despite their limitations V_T -based models have been successfully used for much circuit design work. BSIM4 and MOS Model 9 are modern versions of threshold voltage-based models.

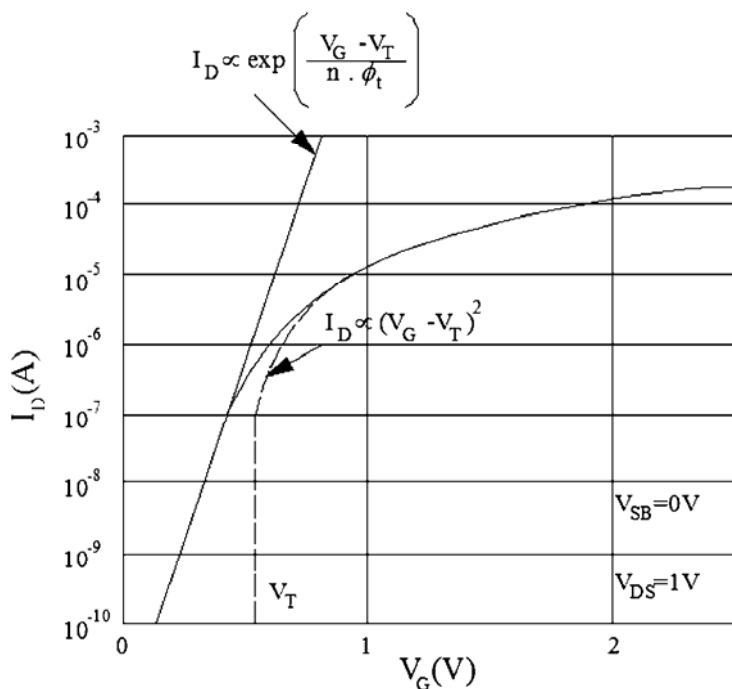


Fig. 3.2 Drain current characteristics given by the strong and weak inversion models. (After [21].)

Finally, in the 1980s charge control models were introduced [22], [23], [24], featuring compact one-equation all-region models for the current and small signal parameters.

Next generation MOSFET models can be divided into two groups based on the adopted key variable. One modeling approach is to solve for the surface potential at the two ends of the channel. The terminal charges, currents and derivatives are then calculated from source and drain surface potentials. These models are called surface potential- or ϕ_s -based models. Examples of such models are SP, MOS model 11, and HiSIM [15]. Another approach is to calculate the density of the inversion charge at the two ends of the channel and formulate the model outputs in terms of these charge densities. For convenience we use the name “inversion charge” or Q'_I -based models to refer to them. Examples of inversion charge models are ACM, EKV, and BSIM5 [15]. Clearly, ϕ_s -based and Q'_I -based models have a common background, but enough differences exist between them to motivate model developers to support one or the other approach. In this chapter we develop the theories and compare results for ϕ_s -based and Q'_I -based core models using the Pao-Sah model as a “golden” reference.

3.2 MOSFET operation

The basic structure of an n-type MOSFET is shown in Fig. 3.3. The p-type doped silicon region, commonly referred to as the bulk or substrate, is connected via the bulk contact. Two heavily *n*-type doped regions, called the source and the drain, are formed in the substrate on either side of the gate. The structure is symmetrical; one cannot distinguish between the source and the drain of an unbiased device. The gate overlaps slightly with the source and drain regions. The region between the source and drain junctions is called the channel region, which has a length L (in the *y*-direction) and a width W (in the *z*-direction). The MOSFET is a four terminal device, the terminals being designated gate, bulk, source and drain.

Under normal operation the source-substrate and drain-substrate junctions are reverse- or zero-biased, and the terminal identified as source, in the case of an n-channel MOSFET, has the lowest reverse bias.

The transistor is said to be off if the gate potential is such that there is no inversion layer. When the gate-to-source potential is positive enough for the formation of an inversion layer, the inversion charge is contacted via the source and drain regions, and a channel current will flow through the device when a potential difference V_{DS} is applied between drain and source. Since the inversion charge depends heavily on the gate potential, the gate can be used to control the current through the channel. As shown in Fig. 3.4, an understanding of the MOS transistor operation requires at least a bidimensional analysis. Fig. 3.4 shows the band bending for an idealized transistor in which all the characteristics are equal following the z -direction as long as narrow-channel effects are negligible. As seen

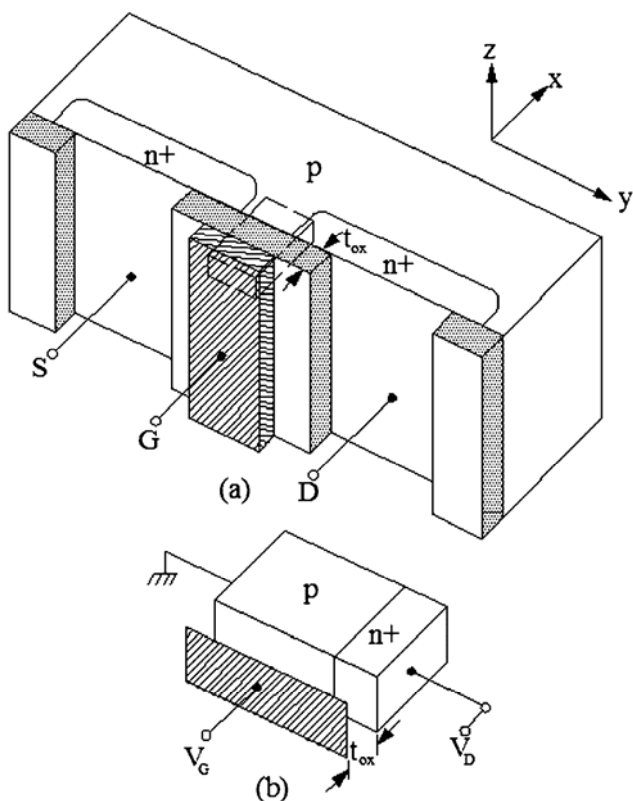


Fig. 3.3 Idealized MOSFET structure. (After [25].)

in Fig. 3.4.(b), because the drain-to-source voltage is higher than zero, the channel inversion level is maximal at the source and decreases toward the drain.

3.3 The Pao-Sah exact I-V model [5], [26]

As shown in Fig. 3.3 and Fig. 3.4, the MOSFET is inherently a two-dimensional (2-D) structure. The input voltage is applied in the x-direction perpendicular to the semiconductor surface in order to modulate the current which flows near the surface in the y-direction when a voltage is applied between source and drain.

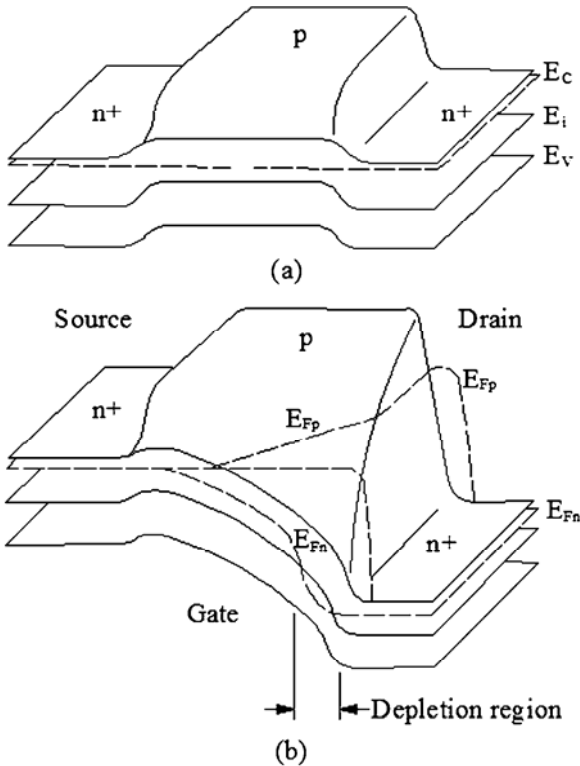


Fig. 3.4 Two-dimensional energy band diagram of a MOSFET operating in (a) linear region (b) saturation. (After [13].)

The first step in creating a compact model is the decomposition of the 2-D problem into two 1-D problems [28]. For a long-channel device, the *gradual channel approximation* is valid, i.e., the longitudinal (y -direction) component of the electric field can be assumed to be much smaller than the transversal (x -direction) component. Thus, the 1-D electrostatic solution is valid for each cross-section of the channel (Fig. 3.5).

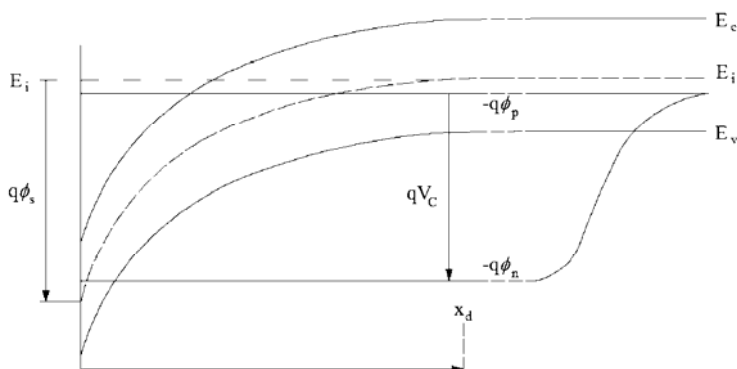


Fig. 3.5 Unidimensional energy band diagram of a cross-section of a MOSFET. (After [7].)

The 1-D x -solution is known as the input voltage equation, which relates applied gate voltage to the electric conditions of the semiconductor surface. We have previously solved the input equation in Section 2.4. The 1-D y -solution is known as the output current equation, which relates the current passing through drain and source to the x -solution and the voltages applied to the source and drain (Fig. 3.6). This solution involves conduction in the channel by the mechanisms of drift and diffusion. The other assumptions for the calculation of the current I_D that flows from drain to source are:

- The hole current (for n-channel) is negligible. This assumption is quite acceptable since under normal operation conditions the source-substrate and drain-substrate junctions are reverse- or zero-biased.
- Recombination/generation along the channel is negligible.
- The current is laminar, that is, the current flows in the y -direction only. The electron current density considering drift and diffusion is given

in terms of the gradient of the quasi-Fermi level by (2.4.6), which is repeated below for convenience.

$$J_n = -qn\mu_n \frac{d\phi_n}{dy} \quad (3.3.1)$$

or

$$J_n = -qn\mu_n \frac{dV_C}{dy} \quad (3.3.2)$$

because the splitting of the quasi-Fermi levels is

$$\phi_n - \phi_p = V_C, \quad (3.3.3)$$

and ϕ_p is constant along the channel, as explained in Section 2.4.1.

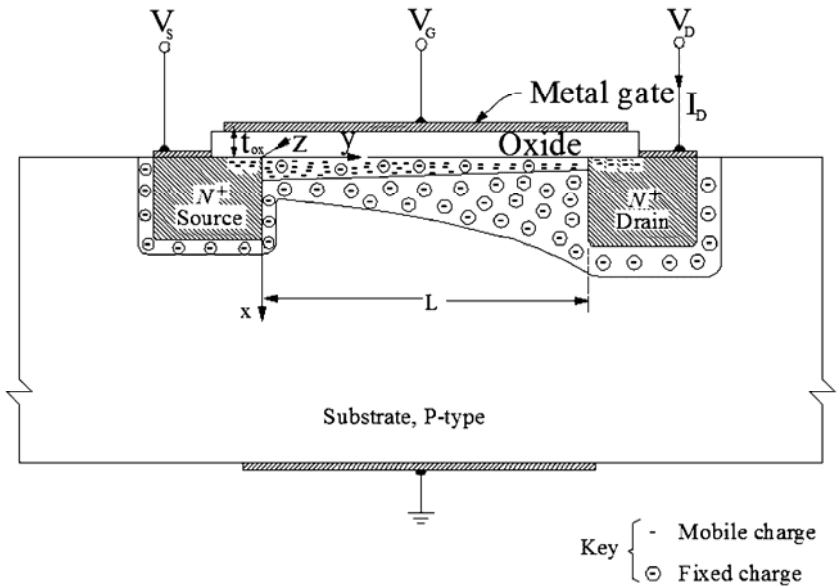


Fig. 3.6 Cross section of a MOSFET. (After [7].)

The channel potential V_C along the channel is such that

$$V_S \leq V_C \leq V_D, \quad (3.3.4)$$

where $V_{S(D)}$ is the source (drain) voltage. Considering steady state ($\partial n / \partial t = \partial p / \partial t = 0$), neglecting generation /recombination ($R_n = G_n = 0$), and assuming laminar flow, the continuity equation for electrons (Appendix D)

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial y} - R_n + G_n \quad (3.3.5)$$

reduces to

$$\frac{\partial J_n}{\partial y} = 0. \quad (3.3.6)$$

The drain current, defined as a positive quantity, is obtained integrating the total current density over the cross-sectional area of the channel

$$I_D = - \int_0^W \int_0^{x_i} J_n dx dz = -W \int_0^{x_i} J_n dx \quad (3.3.7)$$

where W is the transistor width and x_i is taken as the x -coordinate where the electron concentration equals the intrinsic concentration n_i (the bottom end of the inversion channel). Owing to (3.3.6) and (3.3.7), $\partial I_D / \partial y = 0$; thus, the drain current is constant along the channel.

Substituting (3.3.2) into (3.3.7) we obtain

$$I_D = qW \int_0^{x_i} n \mu_n \frac{dV_C}{dy} dx. \quad (3.3.8)$$

Assuming the mobility μ_n to be independent of bias and position, the channel current can be written as [27]

$$I_D = -W \mu_n Q'_I \frac{dV_C}{dy} \quad (3.3.9)$$

because V_C (and dV_C / dy) do not depend on x and, by definition

$$Q'_I = -q \int_0^{x_i} n dx. \quad (3.3.10)$$

Since the current is constant along the channel, the integration of (3.3.9) along the channel, from source to drain, yields

$$I_D = -\frac{\mu_n W}{L} \int_{V_S}^{V_D} Q'_I dV_C, \quad (3.3.11)$$

where L is the channel length. Expression (3.3.11), illustrated in Fig. 3.7, is very general and includes both drift and diffusion mechanisms, thus giving an exact model of the classical (as opposed to quantum-mechanical) long-channel MOSFET.

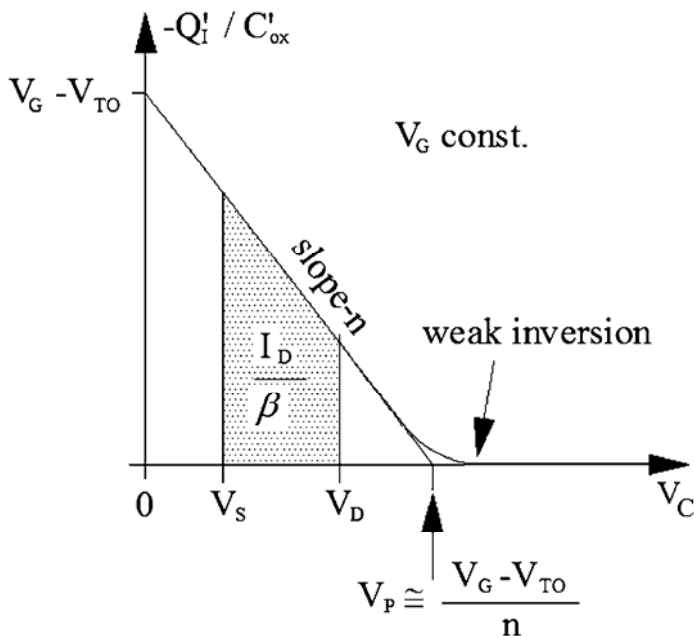


Fig. 3.7 Inversion charge versus the channel potential (splitting of quasi-Fermi potentials), $\beta = \mu_n C'_{ox} W / L$. The drain current is proportional to the shaded surface. (After [29].)

The inversion charge can be calculated in terms of the electric potential by changing the integration variable in (3.3.10) from distance to potential, yielding

$$Q'_I = q \int_{\phi_C}^{\phi_s} \frac{n}{d\phi} d\phi = -q \int_{\phi_C}^{\phi_s} \frac{n}{F} d\phi, \quad (3.3.12)$$

where $\phi_C = \phi(n = n_i) = \phi_F + V_C$ and F is the field function defined in (2.4.17). Substituting (3.3.12) into (3.3.11) we obtain the double integral equation for I_D , referred to as the Pao-Sah model

$$I_D = \mu_n \frac{W}{L} q n_i \int_{V_S}^{V_D} \int_{\phi_C}^{\phi_s} \frac{e^{u-u_F-u_C}}{F} d\phi dV_C. \quad (3.3.13)$$

Equation (3.3.13) is very accurate but it can only be solved numerically. It has been taken as the reference to establish the accuracy of the compact models of the current.

Because there is no general analytically integrable expression for the inversion charge density in terms of the channel potential, the conventional approach has been to develop separate models for strong and weak inversion, where simple expressions, linear and exponential, respectively, for the inversion charge density are available. The complete model is then obtained with the aid of interpolation functions to describe the transition between weak and strong inversion. This approach, which has been widely used to obtain models for electrical simulation, including BSIM4, is no longer considered appropriate for advanced MOS technologies.

To obtain a general (valid from weak to strong inversion) analytical expression for the current, (3.3.11) can be calculated by means of a change of variable. Two alternatives are currently available, namely surface potential models, based on (3.3.14), and inversion charge models, based on (3.3.15)

$$I_D = -\mu_n \frac{W}{L} \int_{\phi_{s0}}^{\phi_{sL}} Q'_I(\phi_s) \frac{dV_C}{d\phi_s} d\phi_s \quad (3.3.14)$$

$$I_D = -\mu_n \frac{W}{L} \int_{Q'_{Is}}^{Q'_{Ip}} Q'_I \frac{dV_C}{dQ'_I} dQ'_I. \quad (3.3.15)$$

In (3.3.14), $\phi_{s0}(\phi_{sL})$ is the surface potential at source (drain) while in (3.3.15), $Q'_{IS(D)}$ is the inversion charge density at source (drain).

3.4 Compact surface potential MOSFET models

To calculate (3.3.14), $Q'_I(\phi_s)$ can be determined from the *charge-sheet approximation*. $Q'_I(\phi_s)$ is given by (2.4.40), which is repeated below for convenience

$$Q'_I = -C'_{ox} (V_G - V_{FB} - \phi_s - \gamma \sqrt{\phi_s - \phi_t}). \quad (3.4.1)$$

$dV_C/d\phi_s$, given by (2.4.39), is also repeated below

$$\frac{dV_C}{d\phi_s} = 1 + \phi_t \frac{2(V_G - V_{FB} - \phi_s) + \gamma^2}{(V_G - V_{FB} - \phi_s)^2 - \gamma^2(\phi_s - \phi_t)}. \quad (3.4.2)$$

As seen in Fig. 3.8, the usual strong inversion expression [7]

$$\phi_s = 2\phi_F + V_C, \quad (3.4.3)$$

which leads to $dV_C/d\phi_s=1$, fails completely in weak inversion and in the saturation region (where the drain side of the channel is in weak inversion). From the definition of the quasi-Fermi level, equation (2.4.6), it is clear that $dV_C=d\phi_s$ implies that the gradient of the quasi-Fermi level is due to the electric field only and, consequently, the diffusion current is neglected. Thus, (3.4.3) does not apply to WI where the main transport mechanism is diffusion.

In WI (including the drain side of a saturated transistor), the channel is almost equipotential as seen in Fig. 3.8. In effect, neglecting the carrier charge (which depends on V_C) in the charge balance equation (2.3.34), gives a surface potential value that depends only on the gate voltage, as in an MOS capacitor. In accordance with the above analysis, the second term in (3.4.2) is relevant in WI and vanishes deep in SI, where $dV_C/d\phi_s \cong 1$. Substituting (3.4.1) and (3.4.2) into (3.3.14) we arrive at

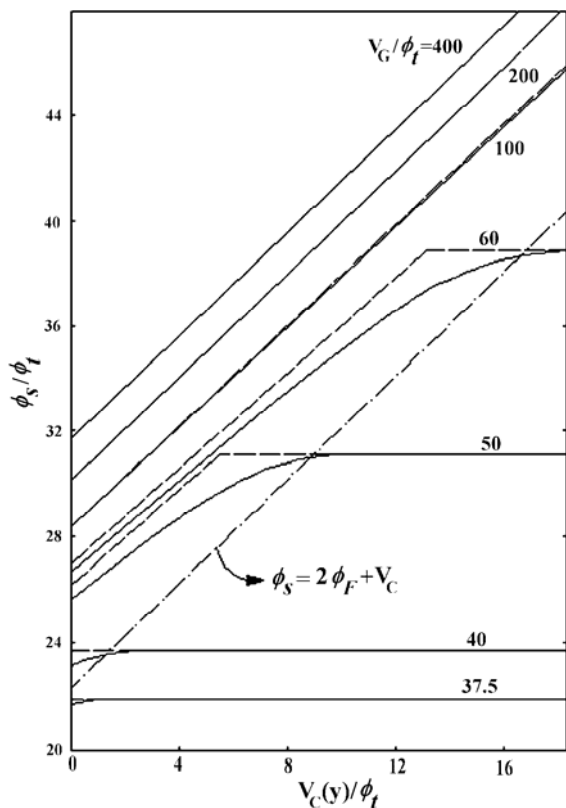


Fig. 3.8 Surface potential vs. channel potential V_C for different gate voltages: (____) exact value, (-----) asymptotic approximation, (-.-.-.-.-) $\phi_s = 2\phi_F + V_C$. (After [18].)

$$\begin{aligned}
 I_D = \mu_n \frac{W}{L} C'_{ox} \int_{\phi_{s0}}^{\phi_{sL}} \left[V_G - V_{FB} - \phi_s - \gamma \sqrt{\phi_s - \phi_t} \right] d\phi_s \\
 + \mu_n \frac{W}{L} C'_{ox} \int_{\phi_{s0}}^{\phi_{sL}} \left[\phi_t \frac{2(V_G - V_{FB} - \phi_s) + \gamma^2}{V_G - V_{FB} - \phi_s + \gamma \sqrt{\phi_s - \phi_t}} \right] d\phi_s.
 \end{aligned} \quad (3.4.4)$$

Equation (3.4.4) allows the derivation of an explicit general expression for the current [16] but the final expression is rather cumbersome. The result can be simplified by noting that the second integrand in the right-hand side of (3.4.4) is relevant for small Q'_i only (weak inversion). This second term corresponds to the diffusion current.

The integrand in the second term of (3.4.4) written in terms of the charge densities is given by

$$\phi_t \frac{2(V_G - V_{FB} - \phi_s) + \gamma^2}{V_G - V_{FB} - \phi_s + \gamma\sqrt{\phi_s - \phi_t}} = \phi_t \left[1 + \frac{Q'_I - \gamma^2 C'_{ox}}{Q'_I + 2Q'_B} \right]. \quad (3.4.5)$$

Since the term in (3.4.5) is relevant for the calculation of the integral in weak inversion only, it can be approximated by setting $Q'_I = 0$. The term in (3.4.5) can then be approximated by

$$\phi_t \left[1 - \frac{\gamma^2 C'_{ox}}{2Q'_B} \right] = \phi_t \left[1 + \frac{\gamma}{2\sqrt{\phi_s - \phi_t}} \right]. \quad (3.4.6)$$

The application of this approximation to (3.4.4) results in:

$$\begin{aligned} I_D = \mu_n \frac{W}{L} C'_{ox} \int_{\phi_{s0}}^{\phi_{sL}} [V_G - V_{FB} - \phi_s - \gamma\sqrt{\phi_s - \phi_t}] d\phi_s \\ + \mu_n \frac{W}{L} C'_{ox} \int_{\phi_{s0}}^{\phi_{sL}} \left[\phi_t \left(1 + \frac{\gamma}{2\sqrt{\phi_s - \phi_t}} \right) \right] d\phi_s \end{aligned} \quad (3.4.7)$$

Integration of (3.4.7) is straightforward, leading to the result known as Brews' charge-sheet model [17]

$$I_D = I_{drift} + I_{diff} \quad (3.4.8)$$

$$\begin{aligned} I_{drift} = \mu_n \frac{W}{L} C'_{ox} \left\{ (V_G - V_{FB})(\phi_{sL} - \phi_{s0}) - \frac{1}{2}(\phi_{sL}^2 - \phi_{s0}^2) \right. \\ \left. - \frac{2}{3} \mathcal{H}(\phi_{sL} - \phi_t)^{3/2} - (\phi_{s0} - \phi_t)^{3/2} \right\} \end{aligned} \quad (3.4.9)$$

$$I_{diff} = \mu_n \frac{W}{L} C'_{ox} \phi_t \left\{ (\phi_{sL} - \phi_{s0}) + \mathcal{H}(\phi_{sL} - \phi_t)^{1/2} - (\phi_{s0} - \phi_t)^{1/2} \right\}. \quad (3.4.10)$$

Fig 3.9 displays plots of both surface potential at source and drain as well as the drain current, decomposed into drift and diffusion components, as functions of the gate voltage. At low gate bias ϕ_{s0} and ϕ_{sL}

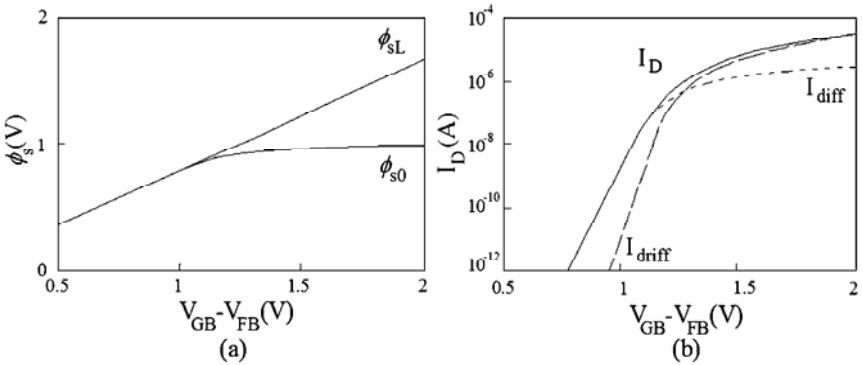


Fig. 3.9 (a) Surface potentials at the source (ϕ_{s0}) and drain (ϕ_{sL}) sides and (b) the corresponding channel current $I_D (=I_{drift}+I_{diff})$ in terms of the effective gate bias $V_{GB}-V_{FB}$ ($V_D=1V$, $V_S=0V$). (After [21].)

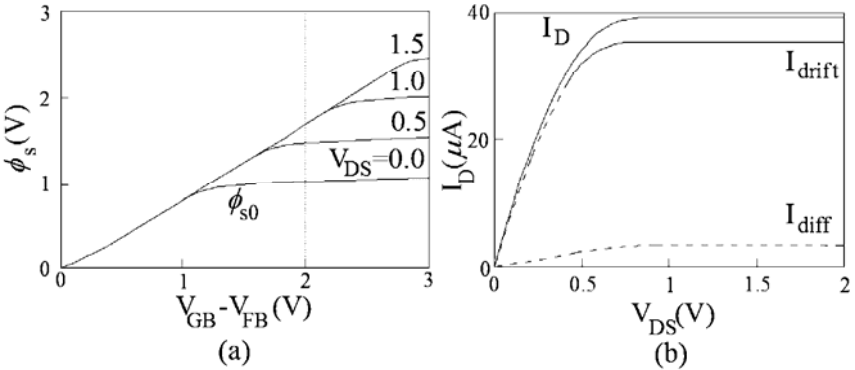


Fig. 3.10 (a) Surface potentials at the source (ϕ_{s0}) and the drain (ϕ_{sL}) sides as functions of of the effective gate bias $V_{GB}-V_{FB}$ for various values of the drain-source voltage V_{DS} and (b) the corresponding channel current $I_D (=I_{drift}+I_{diff})$ as a function of the drain-source voltage V_{DS} ($V_{GB}-V_{FB}=2V$ and $V_S=0V$). (After [21].)

are nearly equal, the channel is equipotential and the diffusion current is the dominant component of the drain current. In this region, the values ϕ_{s0} and ϕ_{sL} must be calculated with high accuracy because the current is calculated as the difference between very close numbers. Approximate explicit expressions for ϕ_s are not acceptable for calculating (3.4.10) in WI. For high gate bias, the source is in strong inversion, and drift prevails over diffusion.

The drain current (in SI) and its components have been plotted as a function of the drain voltage in Fig. 3.10. At low drain-source bias V_{DS} , the drain is also in strong inversion. As a result, $\phi_{sL}-\phi_{s0}$ is approximately equal to V_{DS} and the channel current I_D increases almost linearly with V_{DS} . In this case, the MOSFET acts as a (gate-bias dependent) resistor between drain and source and, as a consequence, this bias region is referred to as the *ohmic* or *linear region*. At high V_{DS} , the drain becomes weakly inverted. The surface potential at the drain ϕ_{sL} is pinned to a value independent of V_{DS} , and as a consequence, I_D also becomes independent of V_{DS} . In other words, I_D saturates for V_{DS} above a certain saturation voltage V_{DSsat} . The drain current expression given by equations (3.4.8), (3.4.9) and (3.4.10) is of the form

$$I_D = \frac{W}{L} \{ (f(\phi_{sL}) - f(\phi_{s0})) \}. \quad (3.4.11)$$

The symmetry of the transistor is clear in equation (3.4.11). If the potential of the source and drain are interchanged, I_D changes sign but its value remains constant as is the case for the current in a linear resistor. The dependence of the current on the geometry is also the same as in a linear resistor (for the long-channel MOSFET analyzed here).

3.4.1 Linearized surface potential models

Most advanced surface potential models [21], [30] introduce an additional approximation, the substitution of the term of the depletion charge density with the first two terms of the Taylor expansion around a certain value of the surface potential ϕ_{se} . In this case, expressions (3.4.9) and (3.4.10) are written as follows

$$I_{drift} = -\frac{W}{L} \mu_n \left[\bar{Q}'_I - n_e C'_{ox} \left(\phi_{se} - \frac{\phi_{sL} + \phi_{s0}}{2} \right) \right] (\phi_{sL} - \phi_{s0}) \quad (3.4.12)$$

$$I_{diff} = \frac{W}{L} \mu_n C'_{ox} n_e \phi_t (\phi_{sL} - \phi_{s0}) \quad (3.4.13)$$

where

$$n_e = 1 + \frac{\gamma}{2\sqrt{\phi_{se} - \phi_t}}, \quad (3.4.14)$$

$$\overline{Q'_I} = -C'_{ox} \left(V_G - V_{FB} - \gamma \sqrt{\phi_{se} - \phi_t} - \phi_{se} \right). \quad (3.4.15)$$

The expression of the drift current in (3.4.12) is simplified by selecting $\phi_{se} = (\phi_{sL} + \phi_{s0})/2$ [30]. The total drain current is now written as

$$I_D = -\frac{W}{L} \mu_n (\overline{Q'_I} - n_e C'_{ox} \phi_t) (\phi_{sL} - \phi_{s0}). \quad (3.4.16)$$

Another virtue of choosing $\phi_{se} = (\phi_{sL} + \phi_{s0})/2$ is that the resulting model is inherently symmetric.

3.4.2 Brews-van de Wiele charge-sheet formula for the current

A general charge-sheet current expression can be deduced using the same kind of approximation applied to Baccarani's formula in the last section. The physical meaning of the approximation is clearly understood using expression (2.4.27) of the derivative $dV_C/d\phi_s$ in terms of the capacitances, instead of the surface potential or charge-based expressions used at the beginning of section 3.4. Equation (2.4.27) is rewritten below

$$\left. \frac{dV_C}{d\phi_s} \right|_{V_G} = 1 + \frac{C'_{ox} + C'_b}{C'_i}. \quad (3.4.17)$$

Note that expression (3.4.17) is more general than the charge-sheet approximation because it has been derived from the general equation (2.4.20) without any further approximations.

Since the second term in the right-hand side of (3.4.17) is relevant for small Q'_I only (weak inversion), we can approximate the inversion capacitance in (3.4.17) [31] by

$$C'_i = -\frac{Q'_I}{\phi_t}, \quad (3.4.18)$$

as previously carried out in section 2.4.6 to derive the UCCM. From Fig. 2.20, it is clear that

$$dQ'_I = C'_i (dV_C - d\phi_s). \quad (3.4.19)$$

Substituting (3.4.18) into (3.4.19) we have

$$dV_C = d\phi_s - \phi_t \frac{dQ'_I}{Q'_I}, \quad (3.4.20)$$

which, substituted into expression (3.3.9) of the current, gives the charge-sheet expression for the current

$$I_D = I_{drift} + I_{diff} = -\mu_n W Q'_I \frac{d\phi_s}{dy} + \mu_n W \phi_t \frac{dQ'_I}{dy}. \quad (3.4.21)$$

From the approximate relationship between Q'_I and ϕ_s given in equation (2.4.41), it follows that, for constant V_G ,

$$dQ'_I = nC'_{ox} d\phi_s. \quad (3.4.22)$$

The substitution of (3.4.22) into (3.4.21) allows the current to be written as a function of the inversion charge density

$$I_D = -\frac{\mu_n W}{nC'_{ox}} (Q'_I - \phi_t nC'_{ox}) \frac{dQ'_I}{dy}. \quad (3.4.23)$$

From (3.4.23), one can readily conclude that the diffusion and the drift components are equal if the local inversion charge density is $-nC'_{ox}\phi_t$, the pinch-off charge defined in (2.4.46). In other words, we have chosen the pinch-off voltage as the channel voltage for which the drift and diffusion components of the current are equal.

3.5 Charge control compact model

In the inversion charge-based model represented by (3.3.15), the term dV_C / dQ'_I is calculated from the unified charge control model (UCCM) written in differential form, Eq. (2.4.57), as repeated below

$$dQ'_I \left(\frac{1}{nC'_{ox}} - \frac{\phi_t}{Q'_I} \right) = dV_C. \quad (3.5.1)$$

Substituting (3.5.1) into (3.3.15) yields [32]

$$I_D = -\mu_n \frac{W}{L} \int_{Q'_{IS}}^{Q'_{IP}} Q'_I \left(\frac{1}{nC'_{ox}} - \frac{\phi_t}{Q'_I} \right) dQ'_I. \quad (3.5.2)$$

Integrating (3.5.2) from source to drain results in

$$I_D = \frac{\mu_n W}{L} \left[\frac{Q'_{IS}{}^2 - Q'_{ID}{}^2}{2nC'_{ox}} - \phi_t (Q'_{IS} - Q'_{ID}) \right]. \quad (3.5.3)$$

In (3.5.3) the quadratic term corresponds to the drift current and the linear term to the diffusion current. The slope factor n in (3.5.3) is given by $n = 1 + \gamma / 2\sqrt{\phi_{sa} - \phi_t}$, with ϕ_{sa} being equal to the surface potential calculated from (2.4.43), assuming the inversion charge to be negligible. An alternative derivation of (3.5.3) results directly from integrating (3.4.23) along the channel [22], [23], [33].

As a final comment on the unicity of the UCCM, let us consider the linear relationship between inversion charge and surface potential in (3.4.22), and the consistency between the expressions of the drain current using either the drift/diffusion charge-sheet equation of (3.4.21) or the gradient of the quasi-Fermi potential of (3.3.9):

$$I_D = \frac{\mu_n W}{nC'_{ox}} (-Q'_I + nC'_{ox}\phi_t) \frac{dQ'_I}{dy} = -\mu_n W Q'_I \frac{dV_C}{dy}. \quad (3.5.4)$$

From (3.5.4) it follows that

$$dQ'_I \left(\frac{1}{nC'_{ox}} - \frac{\phi_t}{Q'_I} \right) = dV_C, \quad (3.5.5)$$

which represents the UCCM in differential form. Some authors [34], [35] have used the drift-diffusion equation in (3.5.4) to obtain UCCM; however, as demonstrated in [31], UCCM can be derived simply by using concepts from basic relationships between charge and potential. In fact, the validity of UCCM does not depend upon transport mechanisms.

3.6 Comparison between models [14]

The numerical calculation of the Pao-Sah double integral (3.3.13) will be used as a baseline for the comparison of drain current models. The inversion charge is solved by finite differences with a non-uniform discrete mesh in x .

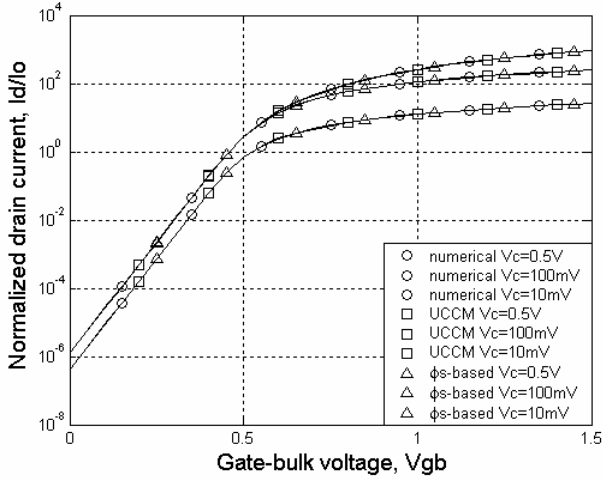


Fig. 3.11 Normalized drain current vs. gate voltage. V_C is the drain voltage and $V_S=0$.

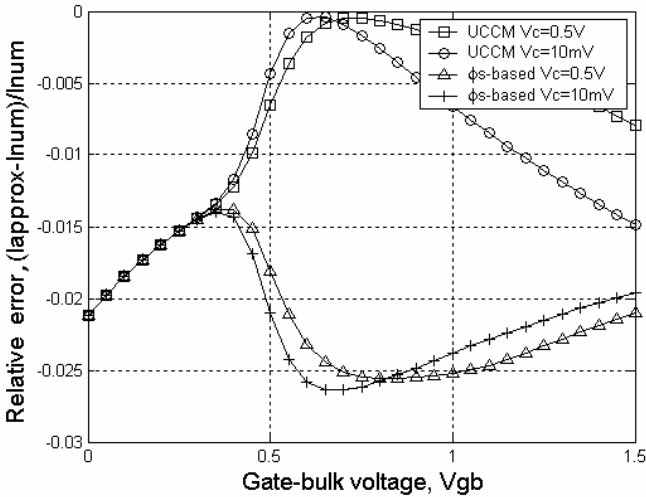


Fig. 3.12 Error in drain current. The Pao-Sah model is used as reference. V_C is the drain voltage and $V_S=0$.

As shown in Fig. 3.11 and Fig. 3.12, the inversion charge- and the ϕ_s -based models give very good approximations for the drain current from weak, through moderate, to strong inversion. Also, note that the

Q'_I - and the ϕ_s -based models are strictly equivalent in weak inversion. An interesting result from Fig. 3.12 is that the charge model gives a better approximation than the charge-sheet surface potential model for moderate inversion while in strong inversion the errors are roughly the same.

For the simulations, the following parameters were used: temperature = 27 °C, oxide thickness = 2 nm, doping concentration = $2 \times 10^{18} \text{ cm}^{-3}$.

3.7 Design-oriented MOSFET model [34], [37]

In this Section we present an MOS transistor model suitable for integrated circuit design. Static and dynamic low-frequency characteristics of the MOSFET are described by single-piece functions with an infinite order of continuity for all regions of operation. Strong and weak inversion are analyzed as asymptotic cases of the single-piece function that describes the overall device behavior. Also, we analyze the long-channel device to give the reader an insight into the definitions we use for design and, in the next chapter, we include the small-geometry effects in the model. Sometimes we refer to this design-oriented model as the current-based model. The design-oriented model has a minimum of physics-based parameters which can be applied to any technology and be useful for statistical analysis.

3.7.1 Forward and reverse components of the drain current

The drain current of a long-channel device can be expressed as the sum of two currents, namely, the forward and reverse saturation currents [34], [36], [37]. For the sake of convenience, we rewrite here expression (3.5.3) for the drain current as

$$I_D = I_F - I_R = I(V_G, V_S) - I(V_G, V_D), \quad (3.7.1)$$

with

$$I_{F(R)} = \mu_n C'_{ox} n \frac{W}{L} \frac{\phi_t^2}{2} \left[\left(\frac{Q'_{IS(D)}}{nC'_{ox}\phi_t} \right)^2 - 2 \frac{Q'_{IS(D)}}{nC'_{ox}\phi_t} \right]. \quad (3.7.2)$$

$I_{F(R)}$ is the forward (reverse) saturation current evaluated at the source (drain) end [34], [36], [37]. Note that, for a long-channel device, the forward (reverse) current depends on both the gate voltage and the source (drain) voltage, being independent of the drain (source) voltage.

Equations (3.7.1) and (3.7.2) emphasize the source-drain symmetry of the MOSFET. Now let us explain how to determine the forward and reverse components of the drain current from the transistor output characteristics, for the example shown in Fig. 3.13 for a long-channel MOSFET. There is a region, usually called the saturation region, where the drain current is almost independent of V_D . This means that, in saturation, $I(V_G, V_D) \ll I(V_G, V_S)$. Therefore, $I(V_G, V_S)$ can be interpreted as the drain current in forward saturation. Similarly, in reverse saturation, I_D is independent of the source voltage. Another interesting property of the MOS transistor arises from its symmetry. Assume, for instance, that V_S in the circuit of Fig. 3.13 is changed to $V_S + \Delta V_S$ and V_G is kept constant. The output characteristics would be the same as in Fig. 3.13 except that the zero of the current would be shifted to $V_S + \Delta V_S$.

Expression (3.7.2) can be rewritten in the form:

$$q'_{IS(D)} = -\frac{Q'_{IS(D)}}{nC'_{ox}\phi_t} = \sqrt{1 + i_{f(r)}} - 1, \quad (3.7.3)$$

where $q'_{IS(D)}$ is the normalized inversion charge density at source (drain) and

$$i_{f(r)} = \frac{I_{F(R)}}{I_S} = \frac{I(V_G, V_{S(D)})}{I_S} \quad (3.7.4)$$

is the forward (reverse) normalized current or inversion coefficient [34], [37], at source (drain) and

$$I_S = \mu_n C'_{ox} n \frac{\phi_t^2}{2} \frac{W}{L} \quad (3.7.5)$$

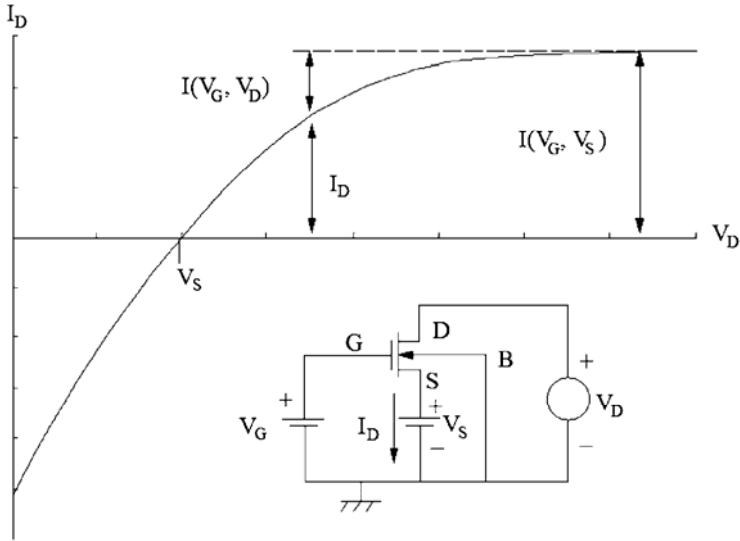


Fig. 3.13 Output characteristics of a long-channel NMOS transistor at constant V_S and V_G . Voltages are referred to the substrate.

is the normalization current, which is four times smaller than the homonym presented in [36]. The factor $I_{SQ} = \mu_n C'_{ox} n \phi_t^2 / 2$, herein denominated the sheet normalization current, is a technological parameter, slightly dependent on V_G , through μ_n and n . As a rule of thumb, values of i_f greater than 100 characterize strong inversion and the transistor operates in weak inversion up to $i_f = 1$. Intermediate values of i_f , from 1 to 100, indicate moderate inversion.

Fig. 3.14 shows the normalization current of a long-channel MOS transistor versus the gate voltage. The variation of the normalization current around its average value is about $\pm 30\%$ for a gate voltage ranging from 0.6V to 5V.

Using the normalized form of UCCM

$$V_P - V_{S(D)} = \phi_t \left[q'_{IS(D)} - 1 + \ln q'_{IS(D)} \right] \quad (3.7.6)$$

and Eq. (3.7.3) we find the following relationship between current and voltage:

$$V_P - V_{S(D)} = \phi_t \left[\sqrt{1 + i_{f(r)}} - 2 + \ln \left(\sqrt{1 + i_{f(r)}} - 1 \right) \right]. \quad (3.7.7)$$

Expression (3.7.7) is a universal relationship for long-channel MOSFETs, valid for any technology, gate voltage, dimensions and temperature.

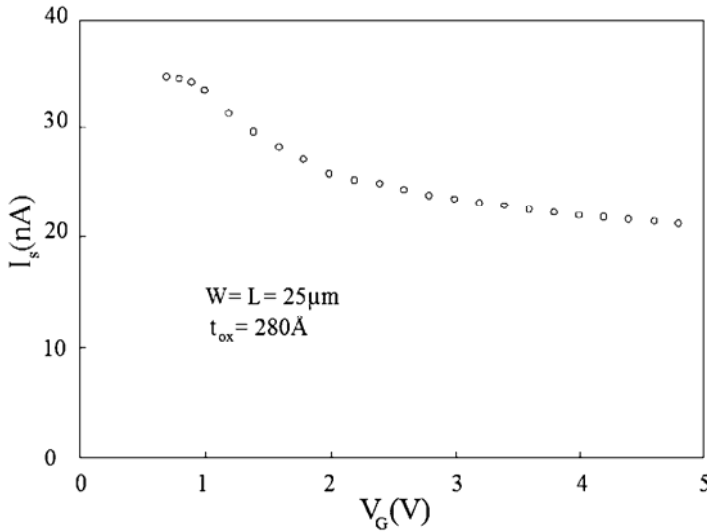


Fig. 3.14 Normalization current of an NMOS transistor ($t_{\text{ox}}=280\text{\AA}$, $W=L=25\mu\text{m}$) vs. gate voltage.

3.7.2 Asymptotic behavior of the drain current in weak and strong inversion

The different modes of operation of the MOS transistor can be defined according to the difference between the values of the source/drain voltage and the pinch-off voltage or, equivalently, according to the ratio of the source/drain inversion charge density to the pinch-off inversion charge density.

Roughly speaking, weak inversion represents a condition for which $-Q'_I < nC'_{\text{ox}}\phi_t$ while, for strong inversion, $-Q'_I \gg nC'_{\text{ox}}\phi_t$. The UCCM expression in (2.4.58) (or the normalized UCCM of (3.7.6)) reduces to

$$\phi_t \ln\left(\frac{Q'_I}{Q'_{IP}}\right) = V_P - V_C + \phi_t \quad (3.7.8)$$

or

$$Q'_I = Q'_{IP} e^{(V_P - V_C + \phi_t)/\phi_t} \quad (3.7.9)$$

for WI, with $V_P - V_C < 0$.

On the other hand, for SI, $V_P - V_C \gg \phi_t$, and UCCM becomes

$$-Q'_I = nC'_{ox} (V_P - V_C). \quad (3.7.10)$$

Expressions (3.7.9) and (3.7.10) are the well-known exponential and linear charge regimes for WI and SI, respectively.

The charge-based current expression (3.5.3) can also be simplified. In WI, the quadratic terms in the charges can be neglected and (3.5.3) reduces to

$$I_D = -\frac{\mu_n W}{L} \phi_t (Q'_{IS} - Q'_{ID}). \quad (3.7.11)$$

Combining (3.7.9) and (3.7.11) yields

$$\begin{aligned} I_D &= I_0 \left[e^{(V_P - V_S)/\phi_t} - e^{(V_P - V_D)/\phi_t} \right] \\ &= I_0 e^{(V_P - V_S)/\phi_t} \left[1 - e^{-V_{DS}/\phi_t} \right], \end{aligned} \quad (3.7.12)$$

where

$$I_0 = \mu_n \frac{W}{L} nC'_{ox} \phi_t^2 e^1. \quad (3.7.13)$$

Expression (3.7.12), which puts into evidence the transistor symmetry, was first introduced in [12].

The drain saturation voltage is the drain voltage at which the reverse current becomes arbitrarily smaller than the forward current. In WI, the current for constant V_G and V_S saturates for $V_{DS} > 4\phi_t$ at a value equal to

$$I_D = I_0 e^{(V_P - V_S)/\phi_t} \cong I_0 e^{(V_G - V_{T0})/n\phi_t} e^{-V_S/\phi_t} \quad (3.7.14)$$

The WI characteristics of a long-channel MOSFET are summarized in Fig. 3.15. The rate of change of the drain current is one decade per $2.3n\phi_t$ of gate voltage variation and one decade per $-2.3\phi_t$ (around -60 mV at 20 °C) of source voltage variation. The output characteristics in weak inversion saturate for a drain-to-source voltage around $4\phi_t$, or around 100 mV at 20 °C.

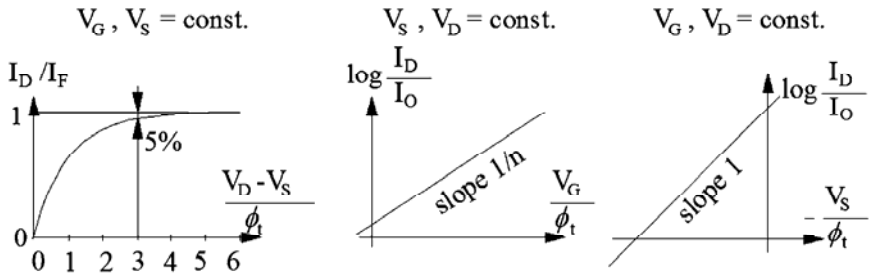


Fig. 3.15 Forward characteristics in WI. (After [29].)

In SI, the linear terms in the charges in (3.5.3) can be neglected and the drain current becomes

$$I_D = \frac{\mu_n W}{L} \left[\frac{Q'_{IS}{}^2 - Q'_{ID}{}^2}{2nC'_{ox}} \right]. \quad (3.7.15)$$

Combining (3.7.10) and (3.7.15) yields

$$I_D = \mu_n C'_{ox} n \frac{W}{2L} \left[(V_P - V_S)^2 - (V_P - V_D)^2 \right] \quad (3.7.16)$$

Equation (3.7.16) [36] shows once again the source-drain symmetry of the transistor, in contrast to classical textbook formulas for strong inversion that do not preserve the structural and, consequently, electrical symmetry of the device. The reader should be advised against using expression (3.7.16) if the terms in parenthesis for an n (p)-channel transistor are less (greater) than zero. If, for example, $V_P - V_D < 0$ for an NMOS transistor and the source is strongly inverted, one calculates the current using (3.7.16) with $V_P - V_D = 0$. In strong inversion, the value of the current tends to saturate when the second term in (3.7.16) becomes much smaller than the first term, *i.e.*, $V_D \geq V_P$ or, equivalently, $V_D \geq (V_G - V_{T0})/n$. In this case, the drain current becomes

$$I_D \cong \mu_n C'_{ox} n \frac{W}{2L} (V_P - V_S)^2 \cong \mu_n C'_{ox} \frac{W}{2nL} (V_G - V_{T0} - nV_S)^2. \quad (3.7.17)$$

$n=1$ (negligible body effect) reduces (3.7.17) to the classical textbook expression for the saturated MOSFET

$$I_D = \mu_n C'_{ox} \frac{W}{2L} (V_G - V_{T0} - V_S)^2 = \mu_n C'_{ox} \frac{W}{2L} (V_{GS} - V_{T0})^2. \quad (3.7.18)$$

3.7.3 Universal dc characteristics

Fig. 3.16 shows the common-gate characteristics of a MOSFET, which are plots of the drain current in saturation versus V_S at constant V_G .

The semi-log plot exhibits a low current region, characterized by a straight line associated with the exponential dependence of the current on the source voltage. For higher currents the strong inversion current follows approximately the parabolic behavior of expression (3.7.18). The curve knee represents the moderate inversion region. Fig. 3.17 displays the common-source characteristics of the same MOSFET employed to obtain the data in Fig. 3.16. As in Fig. 3.16, one identifies the weak inversion region, characterized by the exponential dependence of the drain current on the gate voltage, and the curve knee, which is associated with moderate inversion.

The application of (3.7.6) to the source and drain terminals allows us to eliminate V_p and write the MOSFET output characteristics described by the universal relationship

$$\frac{V_{DS}}{\phi_t} = q'_{IS} - q'_{ID} + \ln \left(\frac{q'_{IS}}{q'_{ID}} \right) \quad (3.7.19)$$

or

$$\frac{V_{DS}}{\phi_t} = \sqrt{1 + i_f} - \sqrt{1 + i_r} + \ln \left(\frac{\sqrt{1 + i_f} - 1}{\sqrt{1 + i_r} - 1} \right). \quad (3.7.20)$$

Expression (3.7.20) demonstrates that the normalized output characteristics of a long-channel MOSFET are independent of technology and transistor dimensions, corroborating again the universality and consistency of the transistor model that has been derived. In Fig. 3.18 we compare the measured output characteristics, for several gate voltages, and the curves obtained by using (3.7.20). Once again, the measured results agree quite well with the theoretical model.

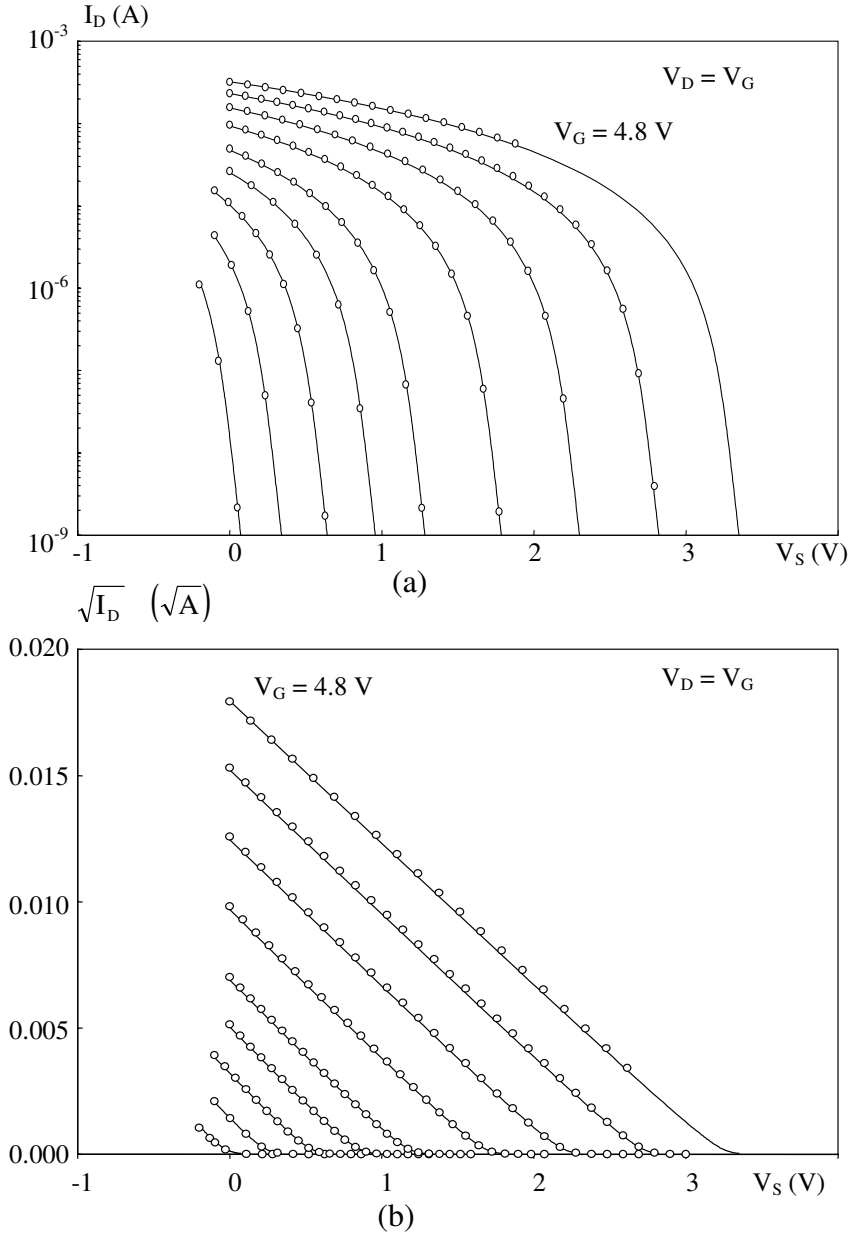


Fig. 3.16 Common-gate characteristics of an NMOS transistor ($t_{ox}=280\text{\AA}$, $W=L=25\mu\text{m}$) in saturation ($V_G=0.8, 1.2, 1.6, 2.0, 2.4, 3.0, 3.6, 4.2$, and 4.8V). (—) simulated and (O) measured data. (a) $\log I_D$ vs. source voltage (b) square root of I_D vs. source voltage.

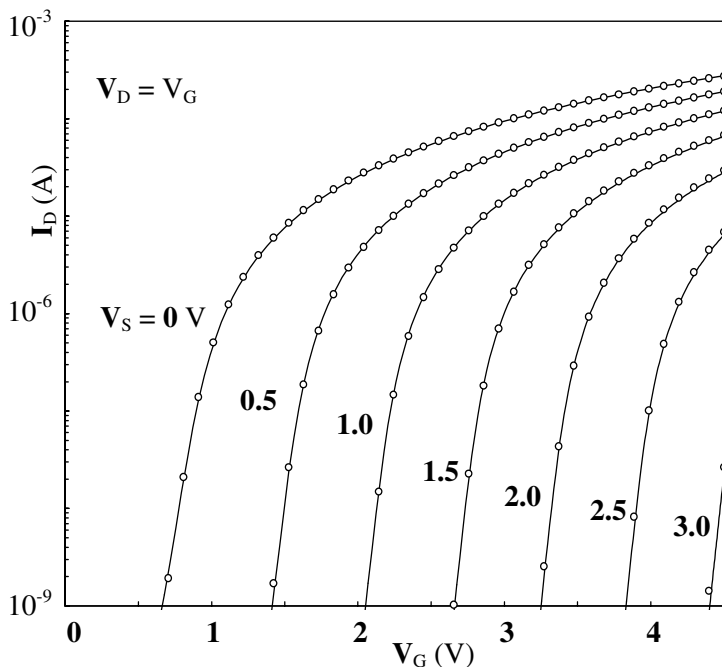


Fig. 3.17 Common-source characteristics of NMOS transistor ($t_{ox}=280\text{\AA}$, $W=L=25\mu\text{m}$) in saturation ($V_S=0$, 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 V). (—) simulated and (o) measured data.

Now, we define V_{DSsat} , which is the value of V_{DS} at which the ratio $q'_{ID}/q'_{IS} = \xi$, where ξ is an arbitrary number much smaller than one. As we shall see, this definition of saturation is extremely practical for long-channel MOSFETs since the ratio of the inversion charge density at the drain over that at the source coincides with the ratio of the transistor conductance at an arbitrary V_{DS} to that determined for $V_{DS}=0$. Note that $(1-\xi)$ represents the saturation level of the MOSFET. If $\xi \rightarrow 1$, the transistor is operating in the linear region, or close to $V_{DS}=0$. If $\xi \rightarrow 0$, the current tends to saturate or, in other words, the saturation level tends to be maximum. From (3.7.19) and $q'_{ID}/q'_{IS} = \xi$, we have

$$V_{DSsat} = \phi_t \left[\ln \left(\frac{1}{\xi} \right) + (1-\xi) \left(\sqrt{1+i_f} - 1 \right) \right]. \quad (3.7.21)$$

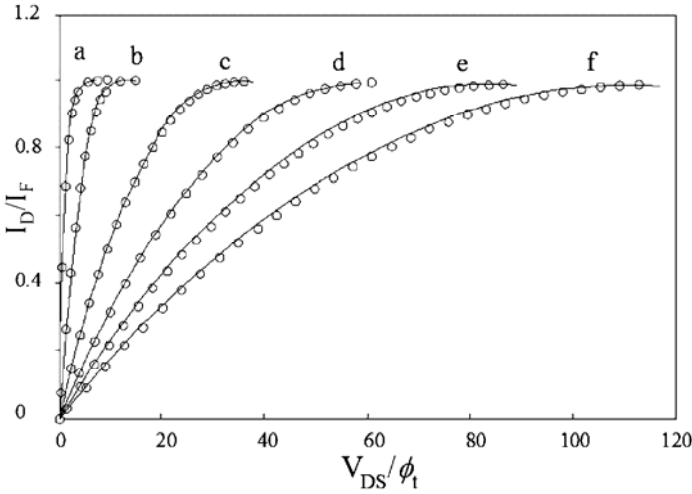


Fig. 3.18 Normalized output characteristics (NMOS transistor, $t_{ox} = 280 \text{ \AA}$ and $W = L = 25 \text{ \mu m}$). I_F has been measured for $V_D = V_G$ and $V_S = 0$. (o): measured data, (—): calculated from . (a) $i_f = 4.5 \times 10^{-2}$ ($V_G = 0.7 \text{ V}$). (b) $i_f = 65$ ($V_G = 1.2 \text{ V}$). (c) $i_f = 9.5 \times 10^2$ ($V_G = 2.0 \text{ V}$). (d) $i_f = 3.1 \times 10^3$ ($V_G = 2.8 \text{ V}$). (e) $i_f = 6.8 \times 10^3$ ($V_G = 3.6 \text{ V}$). (f) $i_f = 1.2 \times 10^4$ ($V_G = 4.4 \text{ V}$).

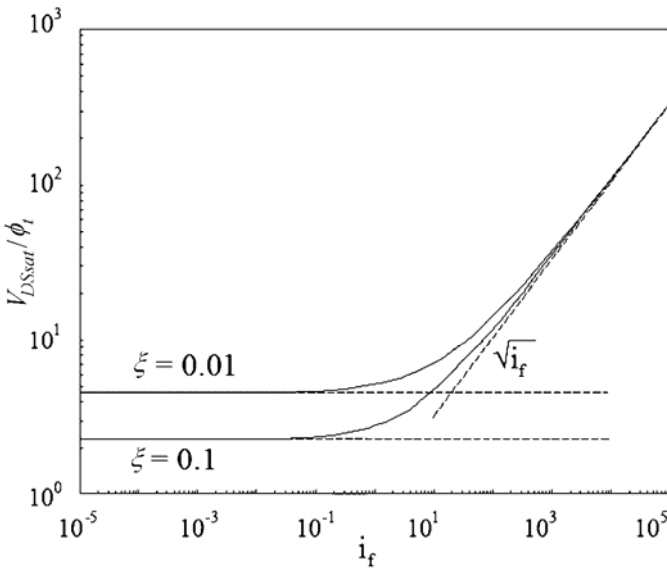


Fig. 3.19 Drain-to-source saturation voltage vs. inversion coefficient, calculated for $\xi = 0.1$ and $\xi = 0.01$.

The definition in (3.7.21) is extremely useful for circuit design since it gives the boundary between the triode and saturation regions in terms of the inversion level. In Fig. 3.19 we present the theoretical drain-to-source saturation voltage for two values of ξ . Note that, in weak inversion, V_{DSsat} (of the order of $4\phi_t$) is independent of the inversion level while in strong inversion it is proportional to the square root of the inversion level. Our definition of saturation is arbitrary but gives designers a very good first order approximation of the lowest V_{DS} required to keep the MOSFET in the “constant current region”.

3.7.4 Small-signal transconductances

In this Subsection we derive a small-signal model for very low-frequency, where we assume that the voltage variations are so slow that the drain current resulting from charge storage variation is negligibly small or, equivalently, the transistor capacitances are not included in the small-signal model.

At low frequencies, the variation of the drain current due to small variations of the gate, source, drain, and bulk voltages is

$$\Delta I_D = g_{mg} \Delta V_G - g_{ms} \Delta V_S + g_{md} \Delta V_D + g_{mb} \Delta V_B, \quad (3.7.22)$$

where

$$g_{mg} = \frac{\partial I_D}{\partial \mathcal{V}_G}, g_{ms} = -\frac{\partial I_D}{\partial \mathcal{V}_S}, g_{md} = \frac{\partial I_D}{\partial \mathcal{V}_D}, g_{mb} = \frac{\partial I_D}{\partial \mathcal{V}_B} \quad (3.7.23)$$

are the gate, source, drain and bulk transconductances, respectively.

If the variation of the gate, source, drain, and bulk voltages is the same, then $\Delta I_D = 0$. Therefore, we can conclude that

$$g_{mg} + g_{md} + g_{mb} = g_{ms}. \quad (3.7.24)$$

Thus, 3 transconductances are enough to characterize the low-frequency small-signal behavior of the MOSFET.

Applying the definitions of source and drain transconductances in (3.7.23) to the equation of the drain current in (3.3.11) results in

$$g_{ms(d)} = -\mu_n \frac{W}{L} Q'_{IS(D)}. \quad (3.7.25)$$

Using (3.7.3), the relationship between inversion charge and inversion level, the source and drain transconductances can be expressed as

$$g_{ms(d)} = \frac{2I_s}{\phi_t} \left(\sqrt{1 + i_{f(r)}} - 1 \right). \quad (3.7.26)$$

The above expression for the transconductance is very useful for circuit design because it is very compact, is valid for any inversion level and uses easily measurable parameters. Moreover, (3.7.26) is a universal relationship for MOSFETs. The only technology-dependent parameter in (3.7.26) is the normalization current, which depends on the transistor aspect ratio as well.

Applying the definition of gate transconductance in (3.7.23) to the relationship $I_D = I_s (i_f - i_r)$ and neglecting the variations of both the slope factor and the mobility with V_G we obtain

$$g_{mg} = I_s \frac{\partial(i_f - i_r)}{\partial V_G}. \quad (3.7.27)$$

From (3.7.7), it follows that:

$$\frac{\partial i_f}{\partial V_p} = -\frac{\partial i_f}{\partial V_s} \quad \frac{\partial i_r}{\partial V_p} = -\frac{\partial i_r}{\partial V_d}. \quad (3.7.28)$$

Recalling that

$$\frac{dV_p}{dV_G} = 1/n, \quad g_{ms} = -I_s \frac{\partial i_f}{\partial V_s}, \text{ and } g_{md} = -I_s \frac{\partial i_r}{\partial V_d}, \quad (3.7.29)$$

it follows that

$$g_{mg} = \frac{g_{ms} - g_{md}}{n}. \quad (3.7.30)$$

(3.7.30) gives the conventional (gate) transconductance in terms of the source and drain transconductances. For a long-channel MOSFET in saturation $i_r \ll i_f$; consequently, $g_{mg} \approx g_{ms}/n$.

Figure 3.20 compares measured and simulated values of both the source and gate transconductances of a long-channel transistor in

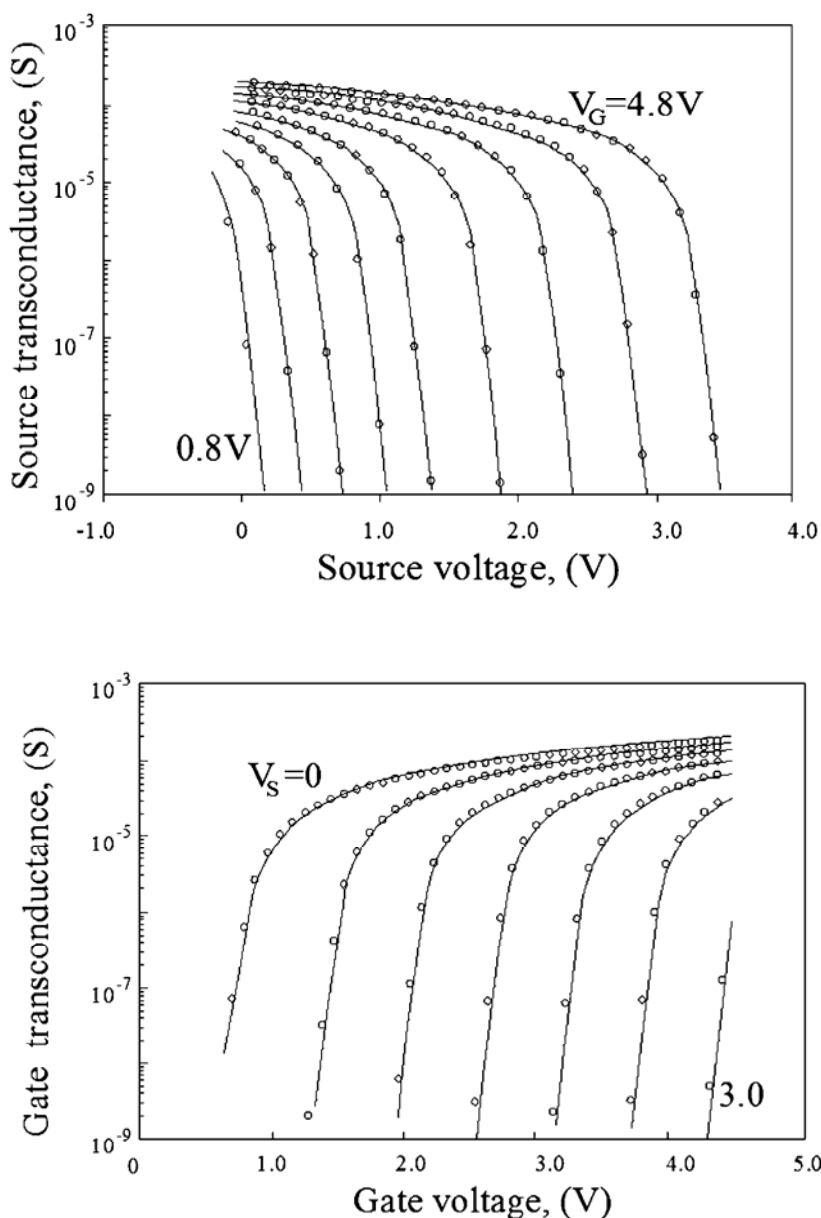


Fig. 3.20 (a) Source transconductance ($V_G=0.8, 1.2, 1.6, 2.0, 2.4, 3.0, 3.6, 4.2$, and 4.8 V) and (b) gate transconductance ($V_S=0, 0.5, 1.0, 1.5, 2.0, 2.5$, and 3.0 V) of an NMOS transistor with $t_{ox}=280$ Å and $W=L=25$ μm. (—) simulated curves; (o) measured curves.

saturation, thus demonstrating the satisfactory accuracy of the proposed model.

It can be observed that all the above calculated transconductances approximate to their well-known asymptotic values in weak and strong inversion [36]. For instance, deep in weak inversion, that is, for $i_f \ll 1$, $\sqrt[3]{1+i_f}$ can be approximated by $1+i_f/2$. Therefore, g_{ms} tends to its expected value, I_F/ϕ . On the other hand, in very strong inversion, g_{ms} is proportional to $\sqrt[3]{I_F}$ since i_f is much greater than one.

3.7.5 The transconductance-to-current ratio

An important design parameter in analog circuits is the transconductance-to-current ratio (g_m/I_D), a measure of the efficiency to translate current (power) into transconductance (speed); thus, speed per unit power consumed [38]. g_m/I_D gives an indication of the inversion level, is strongly related to the performance of a circuit, and provides a tool for calculating transistor dimensions [34], [38]. In the following we will demonstrate that this design parameter can be expressed in terms of a normalized saturation current.

The substitution of I_S by I_F/i_f in (3.7.26) allows one to write the ratio of the source (drain)-transconductance to the forward (reverse) saturation current as

$$\frac{g_{ms(d)}\phi_t}{I_{F(R)}} = \frac{2}{\sqrt{1+i_{f(r)}}+1}. \quad (3.7.31)$$

(3.7.31) is a universal expression for MOS transistors, as is the transconductance-to-current ratio for bipolar transistors. Expression (3.7.31) allows designers to compute the available transconductance-to-current ratio in terms of the inversion level i_f . Moreover, (3.7.31) provides a straightforward procedure, to be seen in Chap. 11, for extracting the value of the normalization current, the most important parameter for current-based design.

The universality of expression (3.7.31) is confirmed in Fig. 3.21 through Fig. 3.23, where measured and simulated transconductance-to-

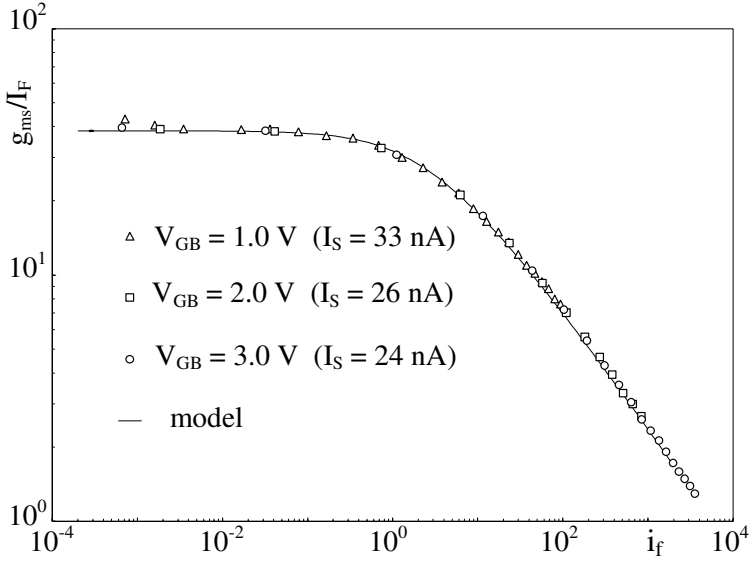


Fig. 3.21 g_{ms}/I_F for different gate voltages. NMOS transistor, $W=L=25\mu\text{m}$, $t_{ox}=280\text{\AA}$.

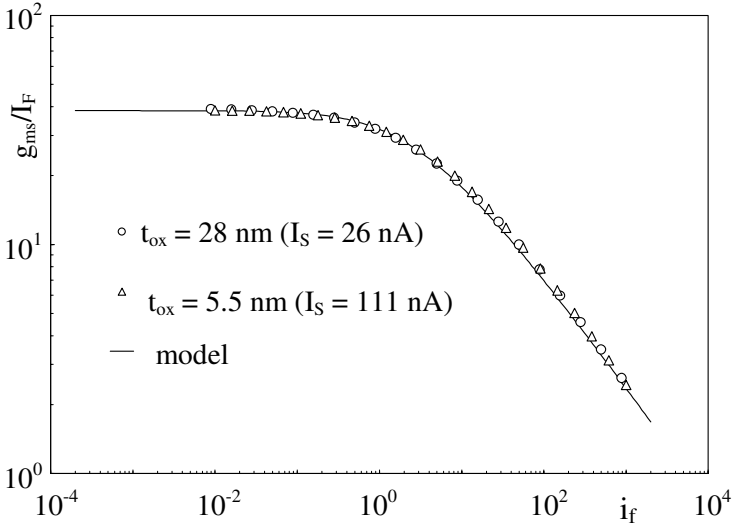


Fig. 3.22 g_{ms}/I_F of NMOS transistors for different technologies. $W=L=25\mu\text{m}$ for $t_{ox}=280\text{\AA}$, $W=25\mu\text{m}$ and $L=20\mu\text{m}$ for $t_{ox}=55\text{\AA}$. $V_{GB} = 2$ V.

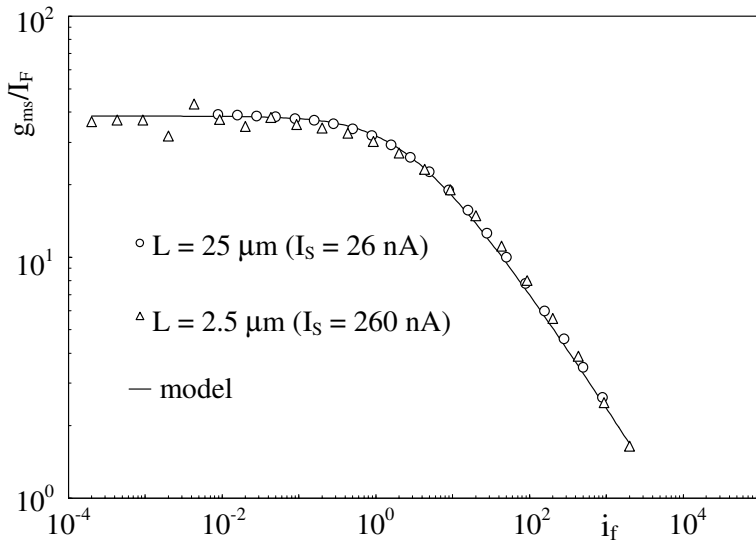


Fig. 3.23 g_{ms}/I_F of NMOS transistors with different channel lengths. $W = 25 \mu\text{m}$, $t_{ox} = 280 \text{ \AA}$, and $V_{GB} = 2 \text{ V}$.

current ratios are plotted for different gate voltages, technologies and channel lengths. All the measurements were taken in saturation, where $I_D \cong I_F$, and are in close agreement with the model we have developed.

3.7.6 Small-signal MOSFET model at low frequencies

In this Section, we derive a small-signal equivalent circuit for the long-channel MOSFET at low frequency. In this derivation, we assume that the voltage variation is so small that the current varies linearly with the voltage. Also, the frequency of the voltage signal is low enough to consider the displacement current associated with the variation of the stored charge to be negligible. The application of KCL to the 4-terminal MOSFET yields

$$i_d + i_s + i_g + i_b = 0. \quad (3.7.32)$$

The symbols in (3.7.32) represent dynamic, small-signal, or ac currents, all of them assumed to enter the transistor. In the model that follows we are going to neglect both gate and bulk currents. This

assumption is acceptable as long as gate leakage is negligible and the p-n junctions are not forward biased. In this case, we have

$$i_d = -i_s. \quad (3.7.33)$$

Now, we can use the transconductances calculated in the previous section to find the drain current. Using small-case symbols to represent dynamic quantities we rewrite (3.7.22) as

$$i_d = g_{mg} v_g - g_{ms} v_s + g_{md} v_d + g_{mb} v_b. \quad (3.7.34)$$

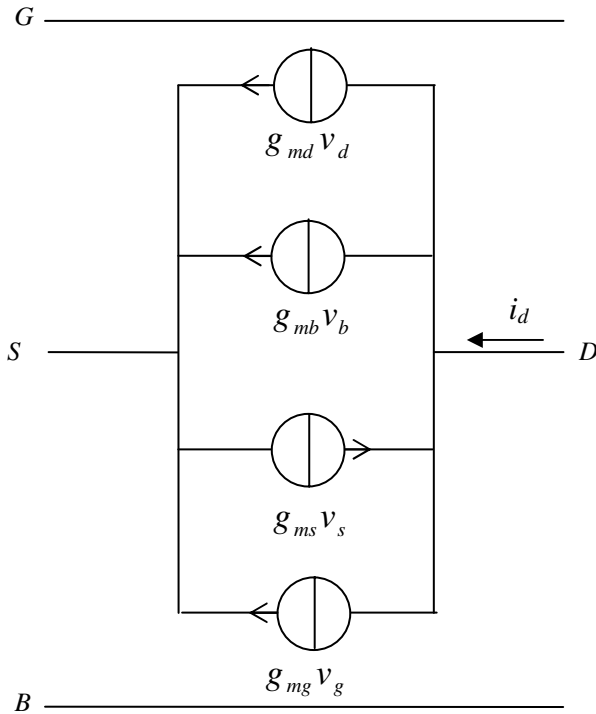


Fig. 3.24 Low-frequency small-signal model of the MOSFET.

The low-frequency small-signal model of the MOSFET that corresponds to expression (3.7.34) is represented in Fig. 3.24. The long-channel transconductances g_{ms} and g_{md} are given by (3.7.31) while g_{mg} is

determined from (3.7.30). The bulk transconductance is calculated from (3.7.24) and (3.7.30), yielding

$$g_{mb} = (n-1)g_{mg} . \quad (3.7.35)$$

The complete medium-frequency small-signal model will be given in Chap. 5, which deals with intrinsic charges and capacitive coefficients of the MOSFET.

Bibliography

- [1] Dawon Kahng, "A historical perspective on the development of MOS transistors and related devices," *IEEE Trans. Electron Devices*, vol. 23, no. 7, pp. 655-657, July 1976.
- [2] Chih-Tang Sah, "Evolution of the MOS transistor – From conception to VLSI," *Proc. IEEE*, vol. 76, no. 10, pp. 1280-1326, Oct. 1988.
- [3] H. K. J. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid-State Electron.*, vol. 7, no. 4, pp. 423-430, Apr. 1964.
- [4] C. T. Sah, "Characteristics of the metal-oxide-semiconductor transistor," *IEEE Trans. Electron Devices*, vol. 11, no. 7, pp. 324-345, July 1964.
- [5] Chih-Tang Sah, "A history of MOS transistor compact modeling," *2005 Nanotechnology Conf.*, presentation slides. [Online]. Available: <http://www.nsti.org/Nanotech2005/WCM2005/#slides>
- [6] Dov Frohman-Bentchowsky and Leslie Vadasz, "Computer-aided design and characterization of digital MOS integrated circuits," *IEEE J. Solid-State Circuits*, vol. 4, no. 2, pp. 57-64, Apr. 1969.
- [7] Richard S. C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New-York, 1970.
- [8] M. B. Barron, "Low level currents in insulated gate field effect transistors," *Solid-State Electron*, vol. 15, no. 3, pp. 293-302, Mar. 1972.
- [9] R. A. Stuart and W. Eccleston, "Leakage currents of MOS devices under surface-depletion conditions," *Electron. Lett.*, vol. 8, no. 9, pp. 225-227, May 1972.
- [10] Richard M. Swanson and James D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE J. Solid-State Circuits*, vol. 7, no. 2, pp. 146-153, Apr. 1972.
- [11] Roger J. Van Overstraeten, G. Declerck, and George L. Broux, "Inadequacy of the classical theory of the MOS transistor operating in weak inversion," *IEEE Trans. on Electron Devices*, vol. 20, no. 12, pp. 1150-1153, Dec. 1973.
- [12] Eric Vittoz and Jean Fellrath, "CMOS analog circuits based on weak inversion operation," *IEEE J. Solid-State Circuits*, vol. 12, no. 3, pp. 224-231, June 1977.
- [13] H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electron.*, vol. 9, no. 10, pp. 927-937, Oct. 1966.

- [14] C. Galup-Montoro, Márcio C. Schneider, and Viriato C. Pahim, "Fundamentals of next generation compact MOSFET models," *Proc. of the 18th Symp. Integr. Circ. Syst. Design*, SBCCI 2005, pp. 32-37.
- [15] Josef Watts, Colin McAndrew, Christian Enz, Carlos Galup-Montoro, Gennady Gildenblat, Chenming Hu, Ronald van Langevelde, Mitiko Miura-Mattausch, Rafael Rios, Chih-Tang Sah, "Advanced compact models for MOSFETs," *Proc. Workshop on Compact Modeling, Nanotech 2005*, pp. 3-12.
- [16] G. Baccarani, M. Rudan, and G. Spadini, "Analytical i.g.f.e.t. model including drift and diffusion currents," *IEE Solid-State and Electron Devices*, vol. 2, no. 2, pp. 62-68, 1978.
- [17] J. R. Brews, "A charge sheet model for the MOSFET," *Solid-State Electronics*, vol.21, no. 2, pp. 345-355, Feb. 1978.
- [18] F. van de Wiele, "A long-channel MOSFET model," *Solid-State Electron.*, vol. 22, no. 12, pp. 991-997, Dec. 1979.
- [19] Daniel P. Foty, *MOSFET Modeling with SPICE – Principles and Practice*, Prentice Hall, Upper Saddle River, NJ, 1997.
- [20] William Liu, *MOSFET Models for Spice Simulation, Including BSIM3v3 and BSIM4*, John Wiley & Sons, New York, 2001.
- [21] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "Physical Background of MOS Model 11," Nat. Lab. Unclassified Report 2003/00239. [Online]. Available: http://www.semiconductors.philips.com/Philips_Models/
- [22] Mary Ann Maher and Carver A. Mead, "A physical charge-controlled model for MOS transistors," in *Advanced Research in VLSI*, P. Losleben (ed.), MIT Press, Cambridge, MA, 1987.
- [23] Mary Ann Maher, *A charge-controlled model for MOS transistors*, PhD Thesis, Caltech, Pasadena, CA, 1989.
- [24] Young Hee Byun, Kwyro Lee, and Michael Shur, "Unified charge control model and subthreshold current in heterostructure field effect transistors," *IEEE Electron Device Lett.*, vol. 11, no. 1, pp. 50-53, Jan. 1990.
- [25] S. M. Sze, *Physics of Semiconductor Devices*, Wiley-Interscience, New York, 1969.
- [26] Narain Arora, *MOSFET Models for VLSI Circuit Simulation Theory and Practice*, Springer-Verlag, Wien, 1993.
- [27] C. T. Sah and H. C. Pao, "The effects of fixed bulk charge on the characteristics of metal-oxide-semiconductor transistors," *IEEE Trans. Electron Devices*, vol. 13, no. 4, pp. 393-409, April 1966.
- [28] W. Shockley, "A unipolar field-effect-transistor," *Proc. IRE*, vol. 40, no. 11, pp. 1365-1376, Nov. 1952.
- [29] Eric Vittoz, "Weak inversion in analog and digital circuits", CCD Workshop (2003), Lund. [Online]. Available: <http://www.es.lth.se/cccd/images/CCCD03-Weak%20inversion-Vittoz.pdf>.
- [30] Gennady Gildenblat, Hailing Wang, Ten-Lon Chen, Xin Gu, and Xiaowen Cai, "SP: An advanced surface-potential-based compact MOSFET model," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1394-1406, Sep. 2004.

- [31] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "Derivation of the unified charge control model and parameter extraction procedure," *Solid-State Electron.*, vol. 43, no. 3, pp. 481-485, Mar. 1999.
- [32] Chan-Kwang Park, Chang-Yeo Lee, Kwyro Lee, Byung-Jong Moon, Young Hee Byun, and Michael Shur, "A unified current-voltage model for long channel n-MOS-FET's," *IEEE Trans. Electron Devices*, vol. 38, no. 2, pp. 399-406, Feb. 1991.
- [33] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electron.*, vol. 38, no 11, pp. 1945-1952, Nov. 1995.
- [34] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510-1519, Oct. 1998.
- [35] Xuemei (Jane) Xi, Jin He, Mohan Dunga, Chung-Hsun Lin, Babak Heydari, Hui Wan, Mansun Chan, Ali M. Niknejad, Chenming Hu, "The development of next generation BSIM for sub-100nm mixed-signal circuit simulation," *Proc. Workshop on Compact Modeling, Nanotech 2004*, pp. 70-73.
- [36] Christian C. Enz, François Krummenacher, and Eric A. Vittoz "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *J. Analog Integr. Circuits and Signal Process.*, vol. 8, pp. 83-114, July 1995.
- [37] Carlos Galup-Montoro, Márcio Cherem Schneider, and Ana Isabela Araújo Cunha, "A current-based MOSFET model for integrated circuit design," Chapter 2 of *Low-Voltage/Low-Power Integrated Circuits and Systems*, pp. 7-55, edited by Edgar Sánchez-Sinencio and Andreas G. Andreou, IEEE Press, Piscataway, NJ, 1999.
- [38] F. Silveira, D. Flandre, and P. G. A. Jespers, "A g_m/I_D based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA," *IEEE J. Solid-State Circuits*, vol. 31, no. 9, pp. 1314-1319, Sep. 1996.

Problems

3.1. a) Using the approximation $dQ'_l = (C'_{ox} + C'_b)d\phi_s = nC'_{ox}d\phi_s$ together with the equation for the drain current, $I_D = \mu W \left(-Q'_l \frac{d\phi_s}{dy} + \phi_t \frac{dQ'_l}{dy} \right)$ show that the drain current of a long-channel device can be written as $I_D = I_F - I_R$ where

$$I_{F(R)} = I(V_G, V_{S(D)}) = I_S \left[\left(\frac{Q'_{IS(D)}}{nC'_{ox}\phi_t} \right)^2 - \frac{2Q'_{IS(D)}}{nC'_{ox}\phi_t} \right] \quad (P3.1.1)$$

with $I_S = (\mu_n C'_{ox} n \phi_t^2 / 2)(W/L)$

b) Determine the approximate relationship between current and charge density for weak inversion and for strong inversion.

c) Using the relationship between transconductance and inversion charge density, deduce the relationship

$$\frac{I_{F(R)}}{\phi_t \cdot g_{ms(d)}} = \frac{\sqrt{1+i_{f(r)}} + 1}{2}. \quad (P3.1.2)$$

Determine approximations of expression (P3.1.2) for the particular cases of weak inversion and strong inversion.

d) Integrate (P3.1.2) to show that the relationship between the normalized forward (reverse) current and the source (drain) voltage is given by

$$\frac{V_{S0} - V_{S(D)}}{\phi_t} = \sqrt{1+i_{f(r)}} - \sqrt{1+i_{f0}} + \ln \left(\frac{\sqrt{1+i_{f(r)}} - 1}{\sqrt{1+i_{f0}} - 1} \right) \quad (P3.1.3)$$

where i_{f0} is the normalized current for $V_S = V_{S0}$. Note that if we choose $V_{S0} = V_P$ (pinch off voltage) then $i_{f0} = 3$ and (P3.1.3) becomes

$$\frac{V_P - V_{S(D)}}{\phi_t} = \sqrt{1+i_{f(r)}} - 2 + \ln \left(\sqrt{1+i_{f(r)}} - 1 \right). \quad (P3.1.4)$$

(P3.1.4) is a modified form of the unified charge control model (UCCM) which can be analogously named as unified current control model (UICM).

e) Using (P3.1.4) show that the normalized drain-to-source voltage can be expressed as

$$\frac{V_{DS}}{\phi_t} = \sqrt{1+i_f} - \sqrt{1+i_r} + \ln \left(\frac{\sqrt{1+i_f} - 1}{\sqrt{1+i_r} - 1} \right). \quad (P3.1.5)$$

f) Assuming that the drain-to-source saturation voltage V_{DSsat} is defined as the value of V_{DS} for which the reverse current is equal to 1% of the forward current, determine the saturation voltage for weak inversion ($i_f < 1$) and strong inversion ($i_f > 1$). Plot V_{DSsat} vs i_f for $0.01 < i_f < 1000$.

g) Assuming that the drain-to-source saturation voltage is defined as the value of V_{DS} for which the ratio $g_{ms}/g_{md}=\xi$, where ξ is much smaller than one, demonstrate that

$$\frac{V_{DSsat}}{\phi_t} = \left(\sqrt{1+i_f} - 1\right)(1-\xi) + \ln\left(\frac{1}{\xi}\right) \cong \left(\sqrt{1+i_f} - 1\right) + \ln\left(\frac{1}{\xi}\right). \quad (\text{P3.1.6})$$

Plot the curves V_{DSsat} vs i_f for $0.01 < i_f < 1000$, with $\xi=0.01$ and $\xi=0.1$ on top of the plot of item f). Also plot the approximate expression $V_{DSsat}/\phi_t \cong \left(\sqrt{1+i_f} - 1\right) + 4$.

3.2. (Reading: C. Galup-Montoro, M. C. Schneider, and I. J. B. Loss, "Series-parallel association of FET's for high gain and high frequency applications," *IEEE J. Solid-State Circuits*, vol. 29, no. 9, pp. 1094-1101, Sep. 1994.)

Transistors can be associated in such a way as to form a trapezoidal shape, with the drain being larger than the source.

a) Consider that ten matched transistors are available. Determine the association of these ten transistors that leads to an equivalent transistor with maximum equivalent channel length. What is the corresponding aspect ratio of the equivalent transistor?

b) Generalize the above result for N matched transistors, assuming that N is an even number.

3.3. Potential profile along the MOSFET channel

The aspect ratio of the transistor equivalent to the series association shown in Fig. P3.3 is W/L . Assume that transistor M_B operates in saturation and that short-channel effects can be neglected. Show that

$$a) \quad I_D = I_{SA}(i_{fA} - i_{rA}) = I_{SB}i_{fB} = I_{SB}i_{rA} \quad (\text{P3.3.1})$$

$$b) \quad i_{rA} = (1 - y/L)i_{fA} \quad (\text{P3.3.2})$$

$$c) \quad V_p = \phi_t \left[\sqrt{1+i_{fA}} - 2 + \ln\left(\sqrt{1+i_{fA}} - 1\right) \right] \quad (\text{P3.3.3})$$

$$d) \quad V_p - V_{ch} = \phi_t \left[\sqrt{1+i_{rA}} - 2 + \ln\left(\sqrt{1+i_{rA}} - 1\right) \right] \quad (\text{P3.3.4})$$

$$e) \frac{V_{ch}(y)}{\phi_t} = \sqrt{1+i_f} - \sqrt{1+\left(1-\frac{y}{L}\right)i_f} + \ln \left(\frac{\sqrt{1+i_f}-1}{\sqrt{1+\left(1-\frac{y}{L}\right)i_f}-1} \right) \quad (\text{P3.3.5})$$

with $i_f = i_{fA}$. Verify that $\frac{V_{ch}(y)}{V_p} \cong 1 - \sqrt{\left(1-\frac{y}{L}\right)}$ in SI and

$$\frac{V_{ch}(y)}{\phi_t} \cong -\ln \left(1 - \frac{y}{L} \right) \text{ in WI.}$$

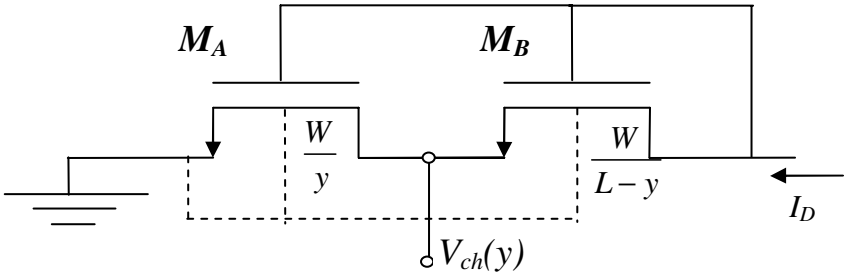


Fig. P3.3 Series association of transistors for computation of the potential profile along the channel.

3.4. (Reading: K. Bult and G. J. G. M. Geelen, “An inherently linear and compact MOST-only current division technique,” *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1730-1735, Dec. 1992).

a) Show that for constant terminal voltages and matched transistors, and neglecting short-channel effects, the following relationship holds for the series association of transistors in Fig. P3.4 (a)

$$\frac{\Delta I_1}{\Delta I_2} = -\frac{W_1/L_1}{W_2/L_2}. \quad (\text{P3.4.1})$$

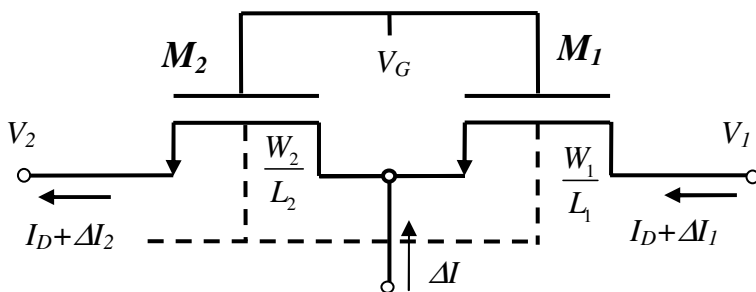
b) Verify that the M-2M network in Fig. P3.4 (b), composed of same-size matched transistors, performs a binary division of the current.

3.5. An n-channel transistor, $W/L=10\text{ }\mu\text{m}/10\text{ }\mu\text{m}$, is implemented in a 0.35 CMOS technology. The sheet normalization current, $I_{SQ}=160\text{ nA}$, and the slope factor $n=1.25$ of the n-channel devices are assumed to be constant over the gate voltage range under consideration in this problem. For your calculations assume that $V_{T0}=0.5\text{ V}$ and $\phi_t=25\text{ mV}$.

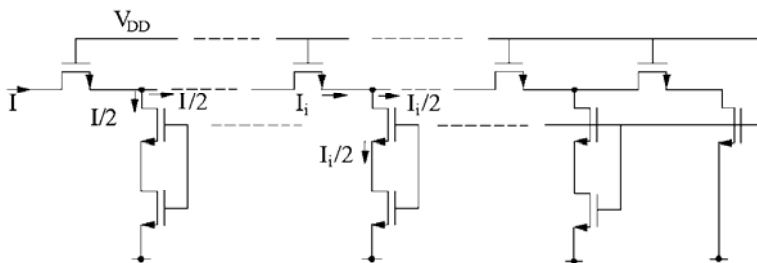
a) Sketch the curve $\log(I_D)$ vs. V_G for $0 \leq V_G \leq 3.3\text{ V}$, $V_S=0$, and $V_D=3.3\text{ V}$. What is the ratio $I_{ON} / I_{OFF} = I_D(V_G = 3.3\text{ V}) / I_D(V_G = 0\text{ V})$?

b) Sketch the curve $\log(I_D)$ vs. V_S for $0 \leq V_S \leq 3.3\text{ V}$, $V_G=V_D=3.3\text{ V}$.

c) Sketch the curve I_D vs. V_D for $V_G=300\text{ mV}$ and $V_S=0$. Determine the forward and reverse components of the drain current for $V_D=0, 50$, and 500 mV .



(a)



(b)

Fig. P3.4 (a) The current division principle , (b) Binary current divider.

This page intentionally left blank

Chapter 4

The Real MOS Transistor: dc Models

In Chapter 3, we considered the MOSFET to be long and wide. We also assumed the mobility to be independent of the electric field. In this chapter we will include the effects of both the mobility variation with the transverse field and velocity saturation on the drain current. We will also consider edge effects and the electrostatic coupling between drain and source. Small-geometry effects account for short and narrow-channel effects on the threshold voltage and drain-induced barrier lowering (DIBL). The common approach to modeling these second-order effects in MOSFETs, also employed here, is to decorate the long-channel model with a list of corrections related to these effects [1]. We conclude the chapter with some benchmark tests concerning compact transistor modeling.

4.1 Effective mobility

The mobility is determined by several scattering mechanisms through which the carriers exchange momentum (and kinetic energy) with the semiconductor. In MOS transistors, the carriers flow near the semiconductor interface with the oxide, and extra scattering mechanisms at the interface lower the mobility of the carriers of the inversion layer (surface mobility) to values of the order of one half of the bulk mobility.

Roughly speaking, the scattering mechanisms are due to the imperfections of the semiconductor crystal, namely lattice vibrations, ionized impurity atoms, and interface related imperfections, such as interface trapped charges and surface roughness. Since lattice vibrations (phonons) depend on the temperature, so does the mobility. The mobility depends also on the doping, and on the electric field component

perpendicular to the current flow (transversal field) which can concentrate the carriers near the surface and subject them to additional scattering. As for the transport in the semiconductor bulk, the velocity of surface carriers saturates for high electric fields in the direction of the current flow (longitudinal field).

In the previous chapter, the mobility was assumed to be constant in order to simplify the calculation of the drain current but in this chapter we will consider mobility dependence (degradation) on both the transversal and longitudinal components of the electric field.

In this section, we assume the MOSFET to be a long-channel device; therefore, the small longitudinal field does not have any appreciable effect on the mobility.

The mobility in the inversion channel has long been a subject of intense investigation [2]. In the particular case of the MOSFET, three mechanisms combine to determine the overall mobility, namely Coulomb scattering, phonon scattering, and surface roughness scattering [3], [4], [5], [6]. These three factors that contribute to the total mobility can be combined using Matthiesen's rule [7], which states that

$$\frac{1}{\mu} = \frac{1}{\mu_{Coul}} + \frac{1}{\mu_{ph}} + \frac{1}{\mu_{sr}}. \quad (4.1.1)$$

In (4.1.1), μ is the total mobility and the factors in the right-hand side of (4.1.1) represent the phenomena contributing to mobility. Figure 4.1 shows the dependence of the inversion layer mobility on the average electric field (to be defined later). Coulomb scattering originates from charged centers, such as the localized charges due to ionized impurities, fixed oxide charges, and interface-state charges. Coulomb scattering is more important for low electric fields, becoming less effective for higher fields due to carrier screening. Phonon scattering is caused by the interaction of carriers with lattice vibrations. Increasing temperatures make the carrier-phonon interaction more intense, thus decreasing the mobility component due to phonon scattering. The third factor affecting the mobility is surface roughness scattering, which displays a strong dependence on the effective field. Strong fields pull carriers toward the surface, making surface roughness the dominant scattering contributing to mobility degradation with strong fields.

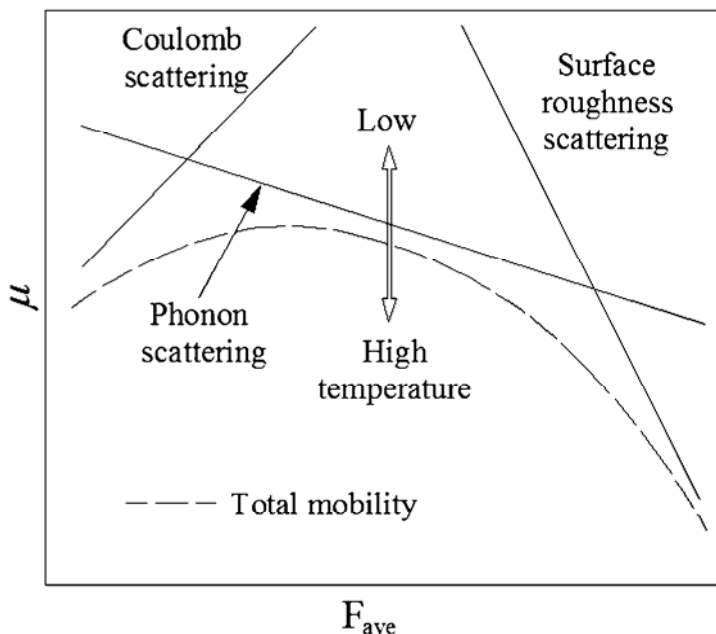


Fig. 4.1 Illustration of the dependence of the mobility in the inversion layer on three dominant scattering mechanisms. (After [4].)

Summarizing, the carrier mobility in MOS transistors depends on temperature, electric field, and doping.

At first glance, mobility seems to be strongly dependent on doping level [8]. However, as demonstrated in [8], when experimental data for mobility were analyzed in terms of an average field in the inversion layer, a universal curve of mobility (almost) independent of doping level corresponded very well to experimental results. After the pioneering work in [8], most of the technical papers on transport in inverted layers use the average field as a characterization parameter. Most of the deviation of experimental data from the universal curve comes from Coulomb scattering. In fact, experimental data show that the deviation from the universal law becomes greater when substrate impurity concentration is high, as shown in [4].

To calculate the average field, we first make use of Gauss' law to calculate the field F within the inversion layer [8], which yields

$$F(x) = -\frac{Q'_B}{\epsilon_s} + \frac{q}{\epsilon_s} \int_x^{x_i} n(x) dx. \quad (4.1.2)$$

When F in (4.1.2) is averaged over the electron concentration in the inversion layer, the following average electric field results

$$F_{ave} = -\frac{Q'_B + \eta Q'_I}{\epsilon_s}. \quad (4.1.3)$$

The calculation of parameter η in (4.1.3) would give 1/2 for a uniform electron concentration along the inversion layer. Experimental data have shown that η depends not only on the type of carrier but also on crystal orientation. It has been found in [9] that η for electron mobility is 1/2 for $\langle 100 \rangle$, and 1/3 for $\langle 110 \rangle$ and $\langle 111 \rangle$ crystals. $\eta=1/3$ for hole mobility for $\langle 100 \rangle$ Si. It has also been found in [9] that μ_{ph} has the same dependence on $F_{ave}^{-0.3}$ for the three crystal orientations. For electrons it was found experimentally [10] that the mobility limited by surface roughness scattering has the following dependence on the effective field

$$\mu_{nsr} \propto F_{ave}^{-2}. \quad (4.1.4)$$

For holes it was found [10] that

$$\mu_{psr} \propto F_{ave}^{-1}. \quad (4.1.5)$$

The reader is referred to [4] and [9] for more accurate expressions of the dependence of mobility on the average field.

The carrier mobility in silicon inversion layers is strongly dependent on temperature [4], [13]. In general, an increase in temperature causes a decrease in mobility. The dependence of mobility on temperature is generally interpreted in terms of the decomposition of mobility according to the Mathiessen's rule, as given in Fig. 4.1. Thus, the dependence of each component of mobility on temperature is studied [4]. It has been experimentally observed that the low-field mobility varies with temperature according to the approximate expression [13]

$$\mu_0(T) = \mu_0(T_0) (T/T_0)^{-m} \quad (4.1.6)$$

for the temperature range 200 – 400 K. T_0 is the reference temperature and m lies in the range 1.2 – 1.4 for p-channel transistors and in the range 1.4 – 1.6 for n-channel transistors [13]. For high fields, the reader is referred to [4], [13] for expressions that can be used in compact MOS models.

Since the mobility depends on the local electric field, F_{ave} , which varies along the channel, we must integrate the variations of the mobility along the channel to obtain a consistent MOSFET model.

Considering here that

$$\mu = \mu(F_{ave}(y)) = \mu(y), \quad (4.1.7)$$

the Pao-Sah current expression (3.3.9) repeated below

$$I_D = -W \mu(y) Q'_I \frac{dV_C}{dy} \quad (4.1.8)$$

can be integrated as

$$I_D \int_0^L \frac{1}{\mu(y)} dy = -W \int_{V_S}^{V_D} Q'_I dV_C. \quad (4.1.9)$$

Defining the effective mobility as

$$\mu_{eff} = \frac{1}{\frac{1}{L} \int_0^L \frac{1}{\mu(y)} dy} = \frac{1}{\frac{1}{L} \int_0^L \left[\frac{1}{\mu_{Coul}} + \frac{1}{\mu_{ph}} + \frac{1}{\mu_{sr}} \right] dy} \quad (4.1.10)$$

the form of the ideal MOSFET expression for the current is preserved in the case of non constant mobility as

$$I_D = -\mu_{eff} \frac{W}{L} \int_{V_S}^{V_D} Q'_I dV_C. \quad (4.1.11)$$

The effective mobility can be calculated in terms of elementary functions considering the empirical expressions for μ_{Coul} , μ_{ph} , and μ_{sr} in terms of the effective field, but the resulting expression is rather cumbersome [10]. A useful approximation is to consider F_{ave} constant along the channel and to calculate its value for an average surface potential.

Another classical approximation [12] assumes that the total mobility depends on two factors, namely bulk mobility and surface mobility, which are combined according to Matthiessen's rule. In addition, the surface mobility is assumed to be inversely proportional to the effective field, yielding the following result [6] for the mobility

$$\mu(y) = \frac{\mu_0}{1 + \alpha_\theta F_{ave}(y)}. \quad (4.1.12)$$

μ_0 , the low-field mobility and α_θ , the scattering constant [13], are considered as fitting parameters [6]. In a MOSFET, the effective field changes along the channel, which implies a corresponding variation in mobility. Substituting (4.1.12) into (4.1.10) it follows that

$$\mu_{eff} = \frac{\mu_o}{\frac{1}{L} \int_0^L \left[1 - \alpha_\theta \left(\frac{Q'_B + \eta Q'_I}{\epsilon_s} \right) \right] dy}. \quad (4.1.13)$$

Since by definition

$$W \int_0^L Q'_B dy = Q_B, \quad W \int_0^L Q'_I dy = Q_I, \quad (4.1.14)$$

where Q_B and Q_I are the total depletion and inversion charges, it follows that

$$\mu_{eff} = \frac{\mu_o}{1 - \alpha_\theta \left(\frac{Q_B + \eta Q_I}{WL\epsilon_s} \right)}. \quad (4.1.15)$$

Explicit expressions for Q_B and Q_I will be given in chapter 5, but the result is rather complicated. For the sake of simplicity, we can use the average value of the field along the channel, as in [14], to define the effective mobility in the channel as

$$\mu_{eff} = \frac{\mu_0}{1 - \alpha_\theta \left(\frac{Q'_{BS} + \eta Q'_{IS}}{2\epsilon_s} + \frac{Q'_{BD} + \eta Q'_{ID}}{2\epsilon_s} \right)}. \quad (4.1.16)$$

The terms in the denominator of (4.1.16) account for the “average” bulk and inversion charges along the channel.

Another simplification considers the effective field to be constant along the channel and equal to its value at pinch-off. Thus

$$\mu_{eff} = \frac{\mu_0}{1 - \alpha_\theta \frac{Q'_{Ba}}{\epsilon_s}}. \quad (4.1.17)$$

In this case, the effective mobility depends only on the depletion pinch-off charge which, in turn, is a function of the gate-to substrate voltage only.

4.2 Velocity saturation

So far, we have assumed the MOSFET to be a long-channel device, *i.e.*, the longitudinal field is assumed to be so low that its effects on device performance are negligible. In most of the practical transistors, however, the effects of the longitudinal field can no longer be neglected. In this section, we examine how the longitudinal field affects carrier transport.

In the previous chapter, we have assumed the drift velocity of carriers v_d to be proportional to the parallel field F_y . In the general case, however, the drift velocity is not proportional to the field. For high fields, the velocity tends to saturate, as Fig. 4.2 shows. This phenomenon is known as velocity saturation.

Much experimental work has been focused on the characterization of mobility in relation to the longitudinal field. One of the most accepted analytical expressions for the drift velocity that fits the experimental results very well [15], [16] is given by

$$v_d = v_{sat} \frac{|F_y| / F_C}{\left[1 + (|F_y| / F_C)^\beta \right]^{1/\beta}}. \quad (4.2.1)$$

In (4.2.1), β and F_C are fitting parameters. β models how rapidly carriers approach the saturation velocity and F_C is the critical field. If

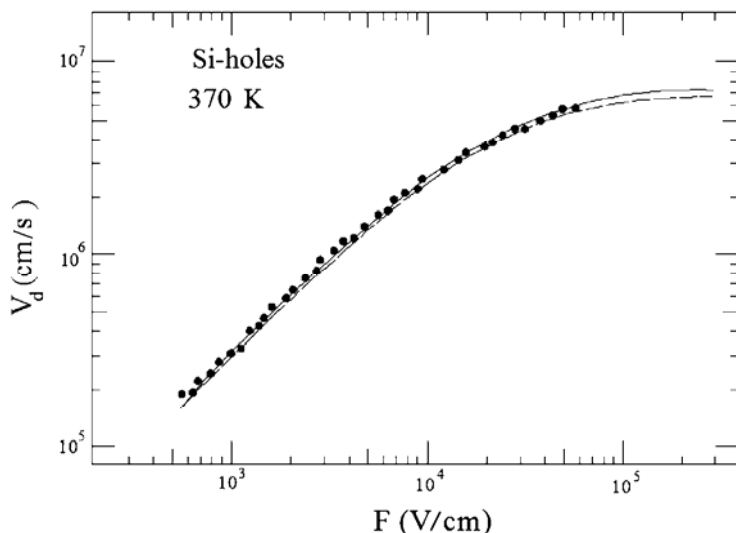


Fig. 4.2 Carrier drift velocity vs. electric field in silicon. (After [15].)

$F_y/F_C \gg 1$, then, from (4.2.1), $v_d = v_{sat}$ is the saturation velocity. For silicon, the saturation velocity in inversion layers is of the order of 5×10^6 to 10^7 cm/s. Typical values of β range between 1 and 2 [15]. Some authors [13], [17] use $\beta=1$ for holes and $\beta=2$ for electrons. We note that for $\beta=1$, the critical field F_C is the value of the electric field at which the carrier velocity is one half of the saturation velocity. For the remaining part of this section we will assume that the drift velocity is given by (4.2.1) with $\beta=1$ for either holes or electrons. Even though this approximation is crude, it will allow us to obtain closed-form solutions for the drain current of an MOS transistor.

We now rewrite the Pao-Sah expression for the drain current to determine the influence of the longitudinal field on I_D

$$I_D = -\mu W Q'_l(y) \frac{dV_C}{dy}, \quad (4.2.2)$$

with the mobility μ derived from (4.2.1) given by

$$\mu = \frac{\mu_s}{1 + |F_y|/F_C}. \quad (4.2.3)$$

In (4.2.3) $\mu_s = v_{sat} / F_C$ is the low-field mobility.

Recalling that the longitudinal field in the charge-sheet model is written in terms of the surface potential and that the surface potential and the inversion charge are linearly related, expression (4.2.2) becomes

$$I_D = - \frac{\mu_s W Q'_I}{1 + \frac{1}{nC'_{ox} F_C} \frac{dQ'_I}{dy}} \frac{dV_C}{dy}. \quad (4.2.4)$$

Now, from UCCM, (2.4.57), we find the relationship between charge and potential, and we rewrite (4.2.4) as

$$I_D = \frac{\mu_s W Q'_I}{1 + \frac{1}{nC'_{ox} F_C} \frac{dQ'_I}{dy}} \frac{dQ'_I}{dy} \left(\frac{-1}{nC'_{ox}} + \frac{\phi_t}{Q'_I} \right). \quad (4.2.5)$$

The integration of (4.2.5) along the transistor channel, from source to drain results in

$$I_D = - \frac{\mu_s W}{nC'_{ox} L} \frac{1}{1 + \frac{Q'_{ID} - Q'_{IS}}{LF_C nC'_{ox}}} \left[\frac{(Q'_{ID} + Q'_{IS})}{2} + Q'_{IP} \right] (Q'_{ID} - Q'_{IS}). \quad (4.2.6)$$

Equation (4.2.6) is a compact expression of the drain current in terms of the inversion charge densities at the drain ends of the channel. It includes the effects of both the transverse and parallel electric fields. One can readily notice that, for low longitudinal fields, (4.2.6) reduces to expression (3.5.3) for the drain current, derived for long-channel devices. Expression (4.2.6) includes both drift and diffusion currents, thus being valid from weak to strong inversion.

The application of UCCM to calculate the denominator of (4.2.6) in strong inversion leads to the following approximation [6] of the drain current when velocity saturation is accounted for

$$I_{D, \text{velocitysaturation}} = \frac{I_{D, \text{novelocitysaturation}}}{1 + \frac{V_{DS}}{LF_C}}. \quad (4.2.7)$$

Note that the result in (4.2.6) is valid for any operating region while (4.2.7) is restricted to strong inversion.

The above calculations are only valid if we consider that the low (longitudinal) field effective mobility is constant along the channel. Expression (4.1.17) for the effective mobility satisfies this condition. However, it is usual in compact modeling to use effective mobility expressions that depend on $V_{D(S)B}$.

In many cases, for theoretical (and numerical) calculations, a normalized expression for the drain current given by (4.2.6) is handy. Normalizing the charges with respect to the pinch-off charge $Q'_{IP} = -nC'_{ox}\phi_t$ and the current with respect to the normalization current

$$I_S = \frac{W}{L} \mu_s n C'_{ox} \frac{\phi_t^2}{2}, \quad (4.2.8)$$

equation (4.2.6) can be rewritten as

$$i_D = \frac{(q'_{IS} + q'_{ID} + 2)}{1 + \zeta (q'_{IS} - q'_{ID})} (q'_{IS} - q'_{ID}), \quad (4.2.9)$$

where the dimensionless short channel parameter ζ ($\zeta \rightarrow 0$ for $L \rightarrow \infty$) is defined as

$$\zeta = \phi_t / LF_C = \phi_t \mu_s / L v_{sat}. \quad (4.2.10)$$

ζ can be viewed as the ratio of the diffusion-related velocity $\mu_s \phi_t / L$ at the source of a saturated transistor over the saturation velocity v_{sat} . Except for the notation, the result in equation (4.2.9) is similar to that presented in [18], [19].

4.3 Saturation charge and saturation voltage

The maximum current that can flow in the channel is limited by the maximum carrier velocity [20]. When electrons at the drain end of the channel reach the saturation velocity, the drain current is expressed as

$$I_{Dsat} = -W v_{sat} Q'_{IDsat}. \quad (4.3.1)$$

In (4.3.1), Q'_{IDsat} is the inversion charge density at the drain end of the channel. Normalizing (4.3.1) in the same way as carried out for the current, expression (4.2.6), yields

$$i_{Dsat} = \frac{I_{Dsat}}{I_S} = \frac{2}{\zeta} q'_{IDsat} . \quad (4.3.2)$$

The saturation condition, relating the source charge to the drain charge in saturation is obtained imposing the equality of the general expression of the drain current with the saturated current. Using the normalized expression for the currents (4.2.9) and (4.3.2), the saturation condition is written as

$$i_{Dsat} = \frac{(q'_{IS} + q'_{IDsat} + 2)}{1 + \zeta (q'_{IS} - q'_{IDsat})} (q'_{IS} - q'_{IDsat}) = \frac{2}{\zeta} q'_{IDsat} , \quad (4.3.3)$$

which is a simple quadratic equation relating the source q'_{IS} and drain saturation q'_{IDsat} charge densities. It is preferable to express q'_{IS} in terms of q'_{IDsat} because the resultant algebraic expression is simpler and better conditioned for numerical calculations. Thus,

$$q'_{IS} = \sqrt{1 + \frac{2}{\zeta} q'_{IDsat}} - 1 + q'_{IDsat} . \quad (4.3.4)$$

Substituting q'_{IDsat} in (4.3.4) in terms of the saturation current using (4.3.2) yields

$$q'_{IS} = \sqrt{1 + i_{Dsat}} - 1 + \frac{\zeta}{2} i_{Dsat} . \quad (4.3.5)$$

We note that, for $\zeta \rightarrow 0$, expression (4.3.5) gives the relationship, (3.7.3), between source inversion charge and forward (saturation) current previously derived for the long-channel transistor. Equation (4.3.5) is a simple physics-based generalization of the long-channel expression, which is very useful for design.

Fig. 4.3 illustrates the dependence of the saturation charge on the source charge, according to expression (4.3.4), for two values of ζ . In weak inversion, the saturation charge density, normalized to the source charge density, tends toward ζ . On the other hand, deep in strong

inversion the channel has the tendency to be homogeneous in saturation when velocity saturation prevails near the source.

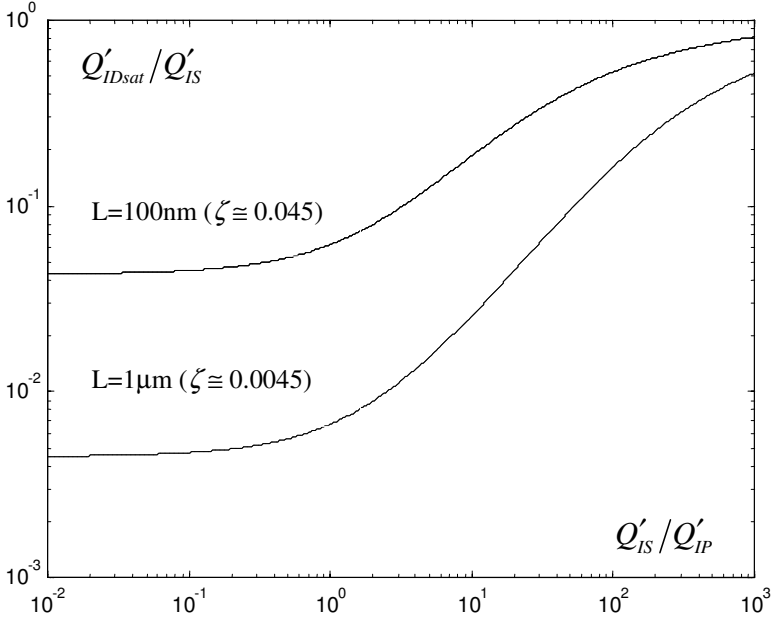


Fig. 4.3 Saturation inversion charge normalized to source inversion charge vs. source inversion charge normalized to pinch-off charge. ($T=290$ K, $F_C=5.5$ V/ μm .)

Fig. 4.4 shows the normalized drain current in saturation in terms of the normalized inversion charge density at the source, according to the expression (4.3.5) for the cases of both an infinitely long-channel ($\zeta=0$) and $\zeta=0.045$. For $\zeta=0.045$, the effect of saturation velocity on the saturation current is almost negligible in weak inversion while the saturation current is reduced by a factor of about four for $q'_{IS}=100$.

The definition of the gate transconductance and expression (4.3.2) together with (4.3.5) yield

$$g_{mg} \cong \frac{2I_s \left(\sqrt{1+i_{Dsat}} - 1 \right)}{n\phi_t} \frac{1}{1 + \zeta \sqrt{i_{Dsat}}} \quad (4.3.6)$$

for practical values of ζ , which are much smaller than 1 (see problem 4.2).

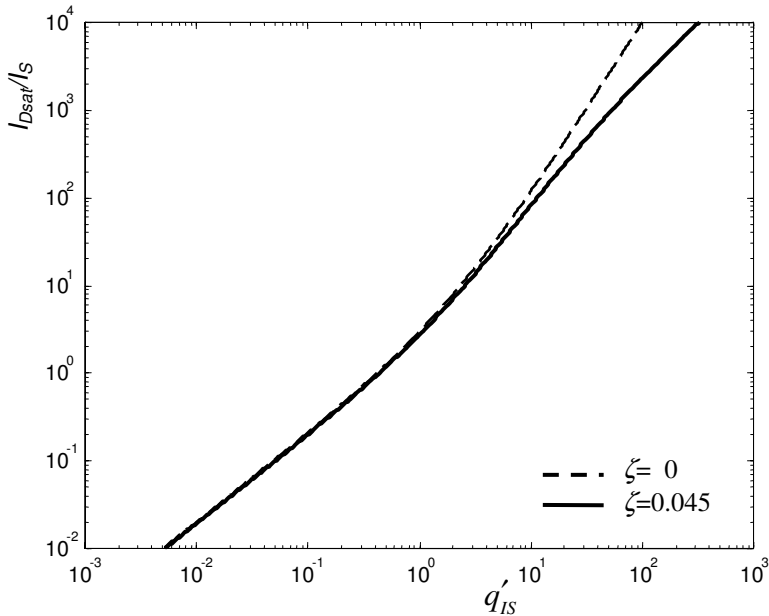


Fig. 4.4 Normalized drain current in saturation vs. normalized inversion charge density at source.

We infer from (4.3.6) that the transconductance of a short-channel device is similar to that of a long-channel device reduced by the last factor in (4.3.6), which accounts for mobility degradation due to the longitudinal field. For normalized currents higher than $1/\zeta^2$, the transconductance increases very slowly with the drain current and reaches a maximum value limited by the saturation velocity.

Fig. 4.5 illustrates the effect of velocity saturation on the transconductance-to-current ratio in saturation, with the channel length as a parameter. The curves were plotted according to expression (4.3.6).

As a final remark, we notice that, for $i_{Dsat} \rightarrow \infty$, the gate transconductance tends towards

$$g_{mg} \cong \frac{2I_s}{n\phi_t\zeta} = WC'_{ox}v_{sat}, \quad (4.3.7)$$

which is the maximum transconductance predicted by conventional strong inversion models. The source transconductance in saturation can be calculated approximately using (4.3.6) with $n=1$.

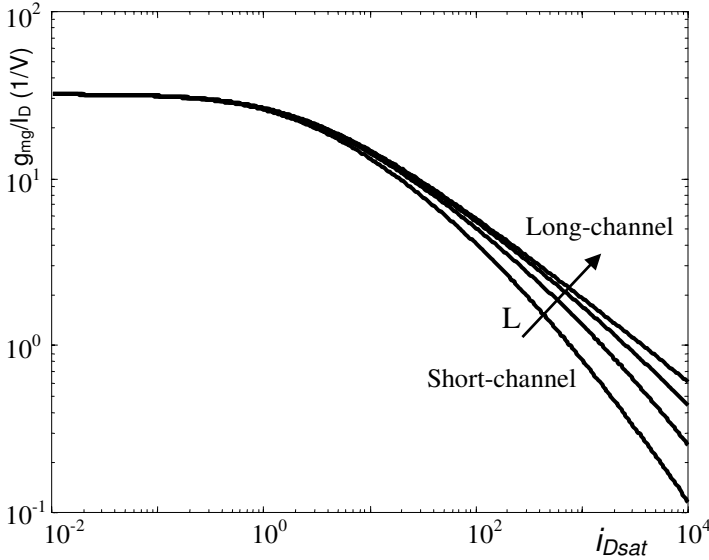


Fig. 4.5 – g_{mg}/I_D for $\zeta = 0.00045$ ($L=10\mu\text{m}$); 0.0045 ($L=1\mu\text{m}$); 0.015 ($L=300\text{nm}$); 0.045 ($L=100\text{nm}$), NMOS transistors in 100 nm CMOS technology.

Now using expression (3.7.19) that gives the drain-to-source voltage in terms of source and drain charge densities yields, for saturation

$$\frac{V_{DSsat}}{\phi_t} = \frac{Q'_{IS} - Q'_{IDSat}}{-nC'_{ox}\phi_t} + \ln \frac{Q'_{IS}}{Q'_{IDSat}}. \quad (4.3.8)$$

Equation (4.3.8), together with expressions (4.3.5) and (4.3.2) relating the source and drain channel charge densities with the saturation current, yields

$$\frac{V_{DSsat}}{\phi_t} = \sqrt{1 + i_{Dsat}} - 1 + \ln \left[1 + \frac{\sqrt{1 + i_{Dsat}} - 1}{\frac{\zeta}{2} i_{Dsat}} \right]. \quad (4.3.9)$$

In weak inversion the saturation voltage becomes

$$\frac{V_{DSsat}}{\phi_t} \cong \ln \left(1 + \frac{1}{\zeta} \right) \cong \ln \left(\frac{1}{\zeta} \right). \quad (4.3.10)$$

As is well known, the saturation voltage is constant in weak inversion. On the other hand, for strong inversion, (4.3.9) is approximated as

$$\frac{V_{DSsat}}{\phi_t} \cong \sqrt{i_{Dsat}} \cong \frac{V_P}{\phi_t} \cong \frac{V_G - V_{T0}}{n\phi_t}, \quad (4.3.11)$$

the classical expression for saturation in strong inversion. Considering that the logarithmic term is only relevant in weak inversion, we can substitute the argument of the logarithm in (4.3.9) for its constant asymptotic value in weak inversion yielding

$$\frac{V_{DSsat}}{\phi_t} = \sqrt{1 + i_{Dsat}} - 1 + \ln \left[\frac{1}{\zeta} \right]. \quad (4.3.12)$$

It is interesting to observe that the physics-based expression above has exactly the same form as the “design-oriented” expression for saturation, given by (3.7.21).

4.4 Channel length modulation

For a MOSFET operating in the saturation regime, the gradual channel approximation becomes less valid, especially in the vicinity of the drain junction, where the two-dimensional nature of the space-charge region must be considered [21].

The usual approach employed to find an analytical formulation for the saturation region divides the channel into two sections. In one of them, closer to the source, the gradual channel approximation is valid while in the other one, closer to the drain, two-dimensional effects must

be accounted for. Using this formulation, the current can be calculated using the expression derived under the gradual channel approximation but considering the effective channel length of the device to be reduced by the length ΔL of the drain section. In addition, the voltage drop in the drain section must be accounted for to calculate the effective drain-to-source voltage V_{DSsat} . The dependence of the effective channel length on the drain-to-source voltage, which is illustrated in Fig. 4.6, is referred to as the channel length modulation (CLM).

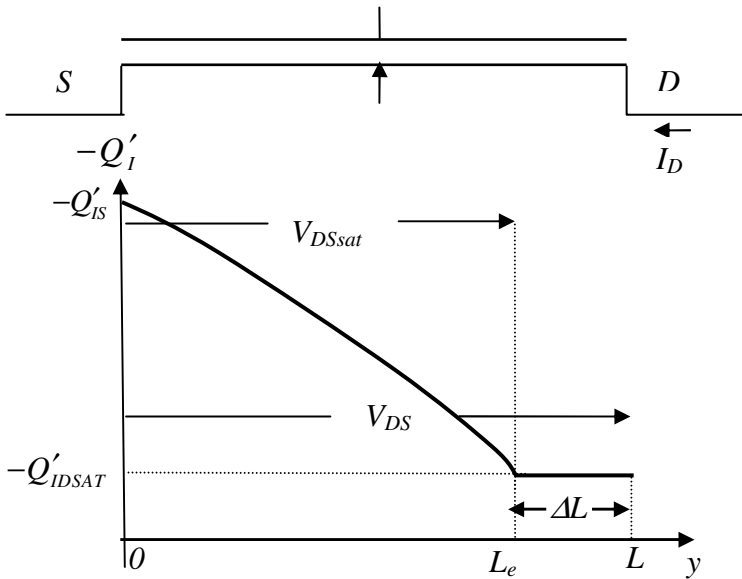


Fig. 4.6 Illustration of the inversion charge density along the channel for a transistor operating in saturation. ΔL represents the channel length shortening.

Some analytical methods [13], [17], [21], [22] are available for the calculation of channel length modulation. An excellent review of them is presented in [13]. In a pseudo two-dimensional formulation, the carrier velocity in the saturated part of the channel is assumed to be equal to the saturation velocity and to flow within the depth of the junction [13]. The channel length shortening ΔL is, according to this model, given by

$$\Delta L = L_C \ln \left[1 + \frac{V_{DS} - V_{DSsat}}{V_E} \right] \quad (4.4.1)$$

if $V_{DS} \geq V_{DSsat}$, and $\Delta L = 0$ if $V_{DS} < V_{DSsat}$. Even though in the derivation of expression (4.4.1) L_C and V_E are physical parameters, they are treated as fitting parameters in compact models [13].

In order to determine the influence of the channel length modulation on the current, we rewrite (4.2.9) as

$$I_D = I_S \frac{1}{1 - \frac{\Delta L}{L}} \frac{(q'_{IS} + q'_{ID} + 2)}{1 + \zeta (q'_{IS} - q'_{ID})} (q'_{IS} - q'_{ID}). \quad (4.4.2)$$

We note that ΔL in (4.4.2) represents the CLM, given by (4.4.1).

4.5 Effect of source and drain series resistances

In the discussion of the dc characteristic so far we have considered the intrinsic transistor, that is, we have neglected the voltage drops in the source and drain diffusion regions. For long-channel devices, the channel resistance is preponderant and the source-drain parasitic resistances can be neglected. For short-channel devices, the source and drain resistances must be considered in strong inversion. The most severe modification of the intrinsic characteristic occurs in the linear region where the channel resistance is minimal. In the saturation region the effect of the drain resistance is minimal because the current is weakly dependent on the drain voltage.

4.6 Ballistic transport

So far, we have considered the drift-diffusion equation for analyzing transport in the MOS transistor. For low electric fields, the average velocity of the carriers is proportional to the field with the mobility as the proportionality factor. The mobility depends on Coulomb scattering, phonon scattering, and surface roughness. For high electric fields, the carrier velocity is no longer proportional to the field and tends to saturate

at around 10^7 cm/s thus imposing a limit on the drain current. As devices get smaller, however, carriers are more likely to traverse the channel without scattering [5], [23]. This would happen if the device channel length were much shorter than the mean free path of electrons of roughly 100 nm [5] for silicon. Under no scattering, carriers injected from the source into the channel would move ballistically.

To give some insight into the ballistic behavior of the MOSFET, let us first assume that carriers in the channel are free from scattering, that all carriers reach the drain without backscattering and that the amount of carriers injected by the drain into the channel is negligible [23]. Therefore, the saturation current can be expressed [23], [24], [25] as

$$I_{Dsat} = -WQ'_{IS} \langle v(0) \rangle, \quad (4.6.1)$$

where $\langle v(0) \rangle$ is the average velocity of the carriers at the beginning of the channel. The maximum value of $\langle v(0) \rangle$ is approximately the equilibrium unidirectional thermal velocity v_T [23], [24], [25]. According to [25], at room temperature $v_T \approx 1.2 \times 10^7$ cm/s for low inversion charge densities but increases for higher charge densities, which leads to velocities greater than the saturation velocity in drift-diffusion transport. In (4.6.1), the carrier charge density Q'_{IS} at the top of the source to channel barrier is fixed by MOS electrostatics in an electrostatically well-designed MOSFET [25]. In a practical device, the backward flow of scattered carriers contributes to reducing the carrier flux and the drain current can be written as

$$I_{Dsat} = -WQ'_{IS} \frac{1-r}{1+r} v_T, \quad (4.6.2)$$

where r is the channel backscattering coefficient [23], [24], [25]. One can calculate the so-called ballistic limit by setting $r = 0$ in (4.6.2). Contrary to drift-diffusion transport, the drain current for pure ballistic transport is independent of the channel length. For zero electric field in the channel, the backscattering coefficient in the channel can be estimated [24] from

$$r = \frac{L}{L + \lambda}, \quad (4.6.3)$$

where L is the channel length and λ is the mean-free-path. When the longitudinal electric field in the channel is not zero, expression (4.6.3) has to be modified [24].

4.7 Short- and narrow-channel effects

Short- and narrow-channel effects are both due to two-dimensional edge effects on the total bulk charge in the silicon surface space-charge region [26]. The short- or the narrow-channel effect becomes important when the channel length or width is comparable with or smaller than the thicknesses of the depletion layer. Short- and narrow-channel effects are often modeled substituting the threshold voltage for an effective threshold voltage which is a function of the channel geometry (length and/or width) and the transistor bias. In ϕ_s -based models, the gate bias is substituted by an effective gate bias that has an effect similar to the effective threshold voltage in V_T -based models. In all these models, the structure of the basic MOSFET model is not altered and, in principle, a specific MOSFET model can accommodate different short/narrow channel effect models.

4.7.1 Short-channel effects

For a sufficiently long and wide n-channel device fabricated within a uniformly doped substrate, the classical definition of V_T is given by equation (2.4.52) rewritten below for convenience as

$$V_{TO} = V_{FB} + 2\phi_F - \frac{Q'_B}{C'_{ox}} \quad (4.7.1)$$

where Q'_B is the depletion charge under the gate for $\phi_s = 2\phi_F$. The threshold voltage V_{TO} for short-channel transistors deviates from that for long-channel devices due to the short-channel effect. For a wide and short-channel device, Q'_B is associated not only with counter charges in the gate electrode but also in the N^+ regions of source and drain. This effect [6], [13], [27], [28] affects decisively the performance of short-channel MOS transistors. In n-channel transistors, the depletion charge

under the gate balanced by charges in the N^+ regions of source and drain depends not only on technological parameters but also on transistor dimensions, and drain and source voltages. Fig. 4.7 shows the dependence of the threshold voltages of n-channel and p-channel devices on the channel length in a 1 μm CMOS technology [29]. The absolute value of the threshold voltage decreases with decreasing channel length and increasing absolute values of V_{DS} . The threshold voltages have been measured for two values of the drain voltage. Note the higher sensitivity of the threshold voltage to the drain voltage for shorter channel length.

The usual approach to include the short-channel effect in the MOSFET model is to replace the threshold voltage with another quantity, denominated the effective threshold voltage [6], [13]. In the charge sharing approach [6], [13], for a uniformly doped substrate, the effective threshold voltage of a short-channel MOSFET is modified according to

$$\Delta V_{Tl} = V_{Tsc} - V_{Tlc} = \frac{Q'_B}{C'_{ox}} (1 - F_l). \quad (4.7.2)$$

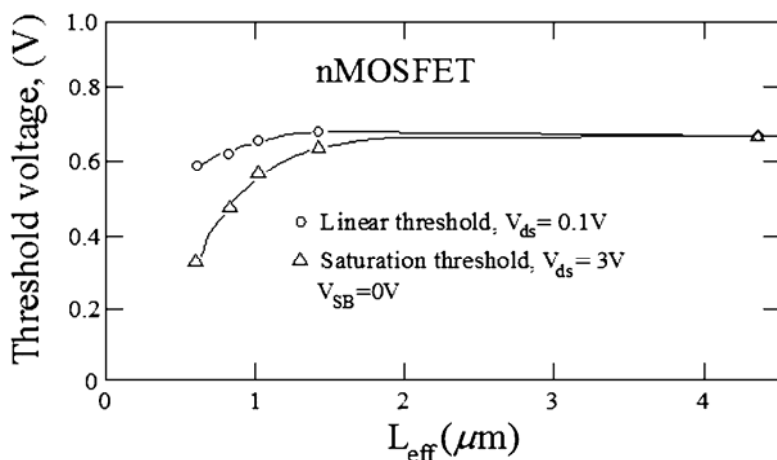
In (4.7.2), the subscripts sc and lc refer to short channel and long channel, respectively, while $F_l (<1)$ is the fraction of the depletion charge controlled by the gate. By using simple geometry and basic physical principles [6], [13], [27], [28], assuming junctions to be cylindrical in shape, and assuming a trapezoidal shape for the gate-controlled depletion charge, it follows that

$$\Delta V_{Tl} = \frac{Q'_B}{C'_{ox}} \frac{X_J}{L} \left(\sqrt{1 + \frac{2X_{dm}}{X_J}} - 1 \right). \quad (4.7.3)$$

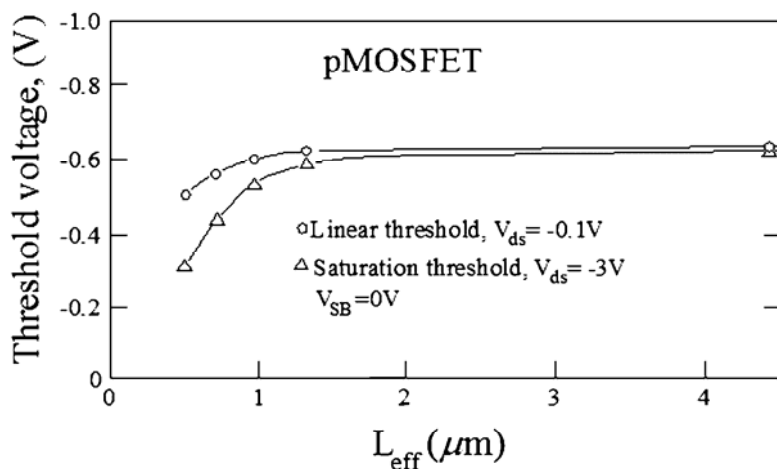
where X_J is the junction depth, and X_{dm} is the depletion depth under the gate. The $1/L$ dependence of the threshold voltage correction is too weak and does not correspond to experimental results.

It should be noted that the division of the depletion charge between the gate and the source and drain is highly idealized. A more realistic model shows an exponential dependence of the ΔV_T term on the channel length. We will give an example of such a model in Section 4.7.4. Another drawback of the charge sharing approach is that the inclusion of the bias dependence in the geometry factor is rather cumbersome. For

these reasons, the charge sharing approach has been abandoned in favor of methods based on approximate solutions of the two-dimensional Poisson equation [30].



(a)



(b)

Fig. 4.7 Short-channel effect for (a) n-channel and (b) p-channel MOSFETs in 1 μm technology. (After [29].)

4.7.2 Reverse short-channel effects

So far, we have assumed the impurity concentration to be uniform along the channel. In most of today's devices, however, the impurity doping along the channel is not uniform and it is often observed that the threshold voltage increases for decreasing channel length. As an illustrative example, Fig. 4.8 shows the threshold voltage vs. channel length in a $0.35\ \mu\text{m}$ CMOS technology. The increase in the threshold voltage for decreasing channel length has been called the reverse short-channel effect and generally originates from higher doping close to the source and drain than in the mid-channel [31]. Current MOSFET technologies make use of some kind of non-uniform channel doping to better control short-channel effects [31]. Simple models of the reverse short-channel effect are available in [32].

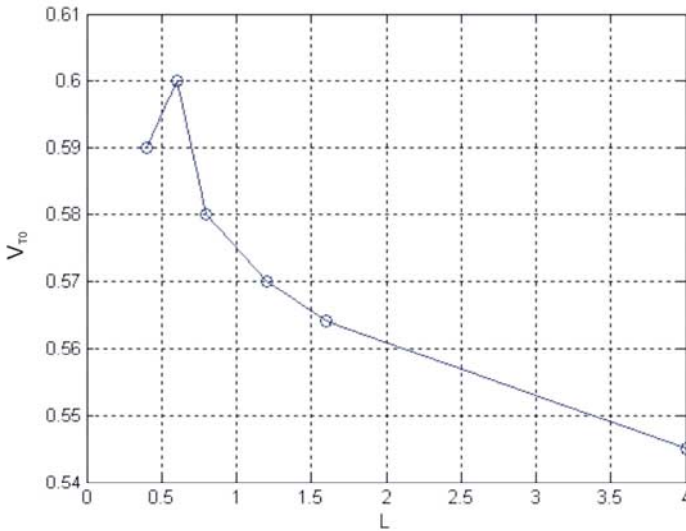


Fig. 4.8 Threshold voltage vs. channel-length for n-MOS transistors in a $0.35\mu\text{m}$ CMOS technology.

4.7.3 Narrow-channel effects

In a MOSFET, the transition from thin oxide to thick oxide is not as sharp as assumed in first order calculations of the threshold voltage; rather, the thin oxide becomes thicker gradually. The cross-section of the MOSFET isolated with the LOCOS (local oxidation of silicon) isolation technique along its width looks like a bird beak, as shown in Fig. 4.9 [34]. As a result of fabrication steps, the polysilicon overlaps with the thick oxide on both sides of the transistor channel. A field implant introduced prior to the formation of the thick oxide prevents the inversion of this region by the overlapped gate.

The charge conservation principle applied to the structure in Fig. 4.9 dictates that

$$Q_G + Q_o + Q_C = 0, \quad (4.7.4)$$

where Q_G , Q_o , and Q_C are the gate charge, fixed oxide charge, and semiconductor charge, respectively. The semiconductor charge is composed of the charge under the thin, tapered, and thick oxide regions. To give an insight into the narrow-channel effect, recall that the threshold voltage is the gate voltage required to induce a certain amount of charge in the semiconductor channel. Let us now assume the transistor

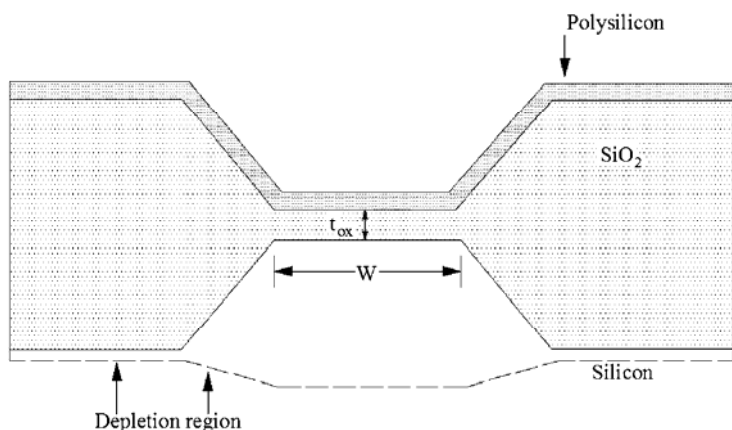


Fig. 4.9 Cross-section of a MOSFET along its width, illustrating the tapered oxide. (After [34].)

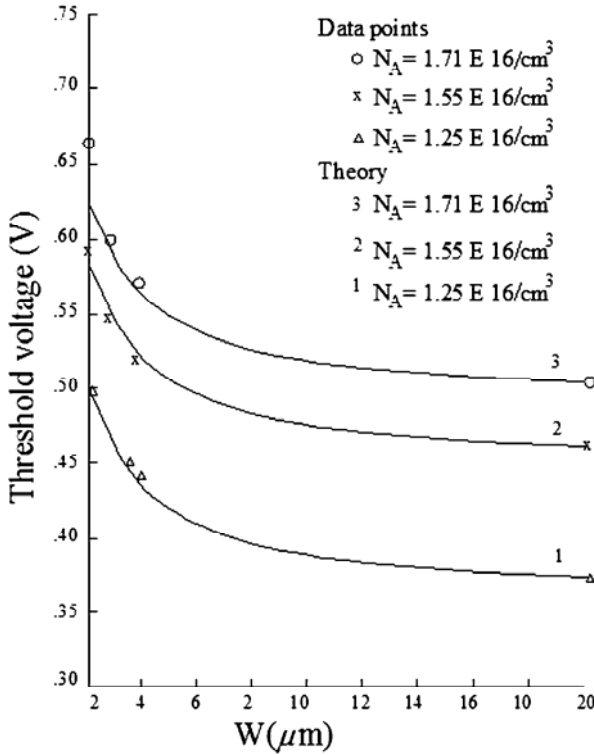


Fig. 4.10 Threshold voltage vs. width for various channel doping values. (After [34].)

to be infinitely wide and the gate voltage to be equal to the threshold voltage. In this case, the amount of charge in the polysilicon gate is counterbalanced by a semiconductor charge in the thin oxide region and the relative amount of charge in the tapered and thick oxide regions is negligible.

Now, assume the gate voltage to be kept constant and W to become narrower. In this case, the gate charge is counterbalanced by both the charge under the thin oxide and the charge under the tapered and thick oxide regions. Therefore, the relative charge under the thin oxide is reduced and, as a result, the gate voltage would have to be increased to restore the threshold charge under the thin oxide. We conclude that the threshold voltage of LOCOS-isolated MOSFETs increases with

decreasing channel width, as illustrated in Fig. 4.10 for various channel doping values.

4.7.4 Drain-induced barrier lowering

An alternative view of the physics of the short-channel effect is the consideration of the potential barrier to channel carriers between the source and drain. Reverse bias of the drain junction creates a field pattern that can lower the potential separating the source from the drain, resulting in increased injection of carriers by the source [13], [35]. This phenomenon is referred to as drain-induced barrier lowering (DIBL).

Fig. 4.11 illustrates the concept of DIBL, as explained in [35]. The device is assumed to operate in the subthreshold region. For the long-channel transistor (curve A), the surface potential is almost constant,

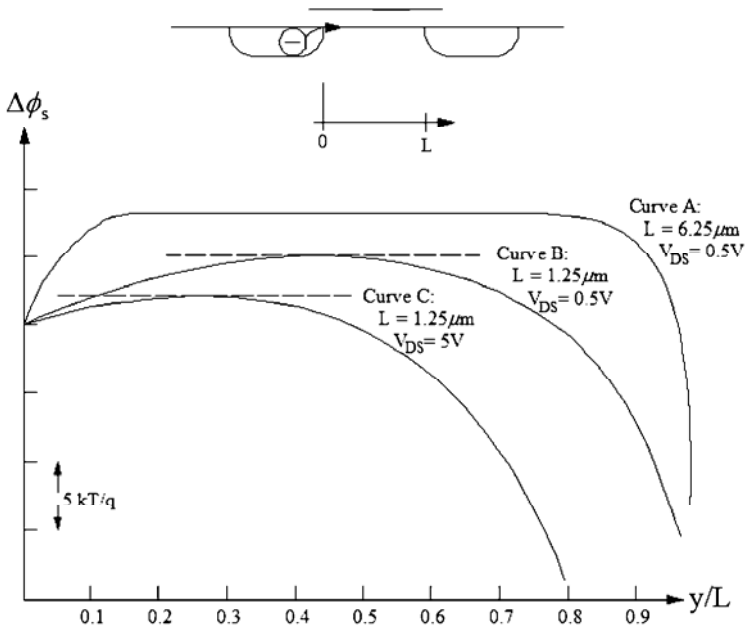


Fig. 4.11 Surface potential distribution for constant gate voltage ($V_G=1.8\text{V}$). Only channel length and drain voltage are varied. (After [35].)

exhibiting a peak which extends over a large portion of the channel. This peak is calculated by solving the one-dimensional Poisson equation in the direction normal to the oxide-semiconductor interface.

For a smaller channel length, the two-dimensional nature of the electric field pattern changes the surface potential profile, as curve B shows. The potential barrier close to the source diminishes, resulting in increased injection of carriers by the source. An increase in drain voltage (curve C), reduces even more the barrier peak, producing a further increase in the current.

Since the early 1970s, a great deal of research work has been done to model DIBL, which is not an easy task owing to its intrinsic two-dimensional nature. As explained in [35] and [36], DIBL results from two effects. One of them is the “proximity” effect, which causes a reduction in the peak potential in the channel, even for $V_{DS}=0$. The other is the “penetration” effect; for short-channel devices at $V_{DS}>0$, the electric field lines penetrate from drain to source, thereby lowering the barrier still more [35]. Although these two effects are not generally separable, they can be in the case of well-designed MOSFETs [36].

The inclusion of the DIBL effect in MOSFET models is generally through the threshold voltage. The approaches employed to model the threshold voltage reduction in short-channel devices use the charge-sharing concept [27], numerical integration of the Poisson equation [37], [38], or some kind of approximation to the solution of the two-dimensional Poisson equation [36], [39], [40]. In general, the charge sharing concept, which has already been explained in Section 4.7.1, does not give accurate results due to the arbitrary division of charge among gate, source, and drain [38]. Numerical analysis is not suitable for circuit analysis since it is associated with a specific device structure and requires substantial computation time.

An analytical model of the dependence of the threshold voltage on technological parameters and bias is shown in [30] as

$$V_T \cong V_{Tc} - \frac{6t_{ox}}{d_1} \left[2(\phi_{bi} - V_{BS}) + V_{DS} \right] \cdot \exp\left(-\frac{\pi L}{4d_1}\right). \quad (4.7.5)$$

Equation (4.7.5) predicts the variation in V_T with L , t_{ox} , and bias. d_1 represents the depletion depth of a long-channel device for a band bending equal to $2\phi_F$. The dependence of V_T on substrate concentration is through $d_1 \propto (N_A)^{-1/2}$. Note that the short-channel effect on the threshold voltage is proportional to the oxide thickness, a result that is consistent with various models [30]. Also note that (4.7.5) does not include an explicit dependence of V_T on junction depth. In order to emphasize the source-drain symmetry of the MOSFET, we rewrite expression (4.7.5) as

$$V_T \cong V_{Tlc} - \sigma \left[(\phi_{bi} + V_{SB}) + (\phi_{bi} + V_{DB}) \right], \quad (4.7.6)$$

where σ is the magnitude of the DIBL factor [41].

4.7.5 Temperature effect on the threshold voltage [13], [42]

The threshold voltage V_{T0} of the MOSFET is given by (2.4.52), repeated below for convenience

$$V_{T0} = V_{FB} + 2\phi_F + \gamma \sqrt{2\phi_F}. \quad (4.7.7)$$

The impact of the temperature on V_{T0} can be found by examining the dependence on the temperature of each term in (4.7.7). We will consider here the case of an n-channel device with a heavily doped n-type polysilicon gate. The analysis for p-channel devices with a p^+ poly-gate is similar.

The flat-band voltage is composed of two terms, the work function difference between the poly-gate and the silicon substrate plus the equivalent oxide charge. For the analysis that follows, let us assume that the contribution of the oxide charge to the flat-band voltage is negligible (which is common in modern VLSI technologies [42]) or, at least, does not change with temperature. In this case,

$$\frac{dV_{FB}}{dT} = \frac{d\phi_{MS}}{dT} = -\frac{d}{dT} \left[\frac{E_g}{2q} + \phi_F \right]. \quad (4.7.8)$$

where E_g is the silicon band-gap energy and $\phi_F = (kT/q) \ln(N_A/n_i)$. Note that function is equal to the contact potential between a heavily doped n-region, for which the Fermi level almost coincides with the

bottom of the conduction band and the p- type substrate. The variation in the intrinsic concentration n_i with temperature is given by

$$n_i = \sqrt{N_c N_v} \exp(-E_g / 2kT) = AT^{3/2} \exp(-E_g / 2kT), \quad (4.7.9)$$

where N_c and N_v are the effective densities of states in the conduction and valence bands. The constant term is $A \approx 4 \times 10^{16} \text{ K}^{-3/2} \text{-cm}^{-3}$ [13]. Considering that γ is independent of the temperature, the following thermal coefficient of the threshold voltage results

$$\frac{dV_{T0}}{dT} = \frac{dV_{FB}}{dT} + \left(2 + \frac{\gamma}{\sqrt{2\phi_F}} \right) \frac{d\phi_F}{dT}. \quad (4.7.10)$$

Now, using the definition of the Fermi level and the expressions for the Fermi level and the intrinsic concentration, we find that

$$\frac{dV_{T0}}{dT} \cong -(2n_0 - 1) \frac{k}{q} \left[\frac{3}{2} + \ln \left(\frac{\sqrt{N_c N_v}}{N_A} \right) \right] + (n_0 - 1) \frac{1}{2q} \frac{dE_g}{dT}, \quad (4.7.11)$$

where $n_0 = 1 + \gamma / (2\sqrt{2\phi_F})$ is the slope factor calculated for $V_G = V_{T0}$.

Let us now assume that $n_0 = 1.25$ and $N_A = 10^{17} \text{ cm}^{-3}$. The thermal coefficient dE_g / dT of the band gap is approximately $-2.73 \times 10^{-4} \text{ eV/K}$ [42]. At $T = 300 \text{ K}$, the thermal coefficient of V_{T0} is approximately -0.82 mV/K . Typical thermal coefficients of the threshold voltage for n-channel devices lie in the range -0.5 to -3 mV/K [6]. Temperature coefficients of p-channel devices are positive. As a consequence, both p-channel and n-channel devices exhibit increasing subthreshold currents as the temperature increases, which implies increased leakage current when the device is in the OFF state ($V_{GS} = 0$).

4.8 Impact of small geometry effects on transistor model

In the case of V_T -based models, the short, narrow, and DIBL effects can be included in the transistor threshold voltage as follows

$$\Delta V_T = V_T - V_{T\infty} \cong \Delta V_{TI} + \Delta V_{Tw} + \Delta V_{TDIBL}. \quad (4.8.1)$$

The three terms in the right-hand side of (4.8.1) represent the short-channel, narrow-channel, and DIBL effects. The subscript ∞ refers to an

infinitely long and wide transistor. As previously mentioned, there are no universal formulas that can be used for the calculation of the threshold voltage deviations. Another simplification of (4.8.1) is the assumption that the overall threshold voltage deviation is the superposition of uncoupled mechanisms. In UCCM, (2.4.58), the pinch-off voltage and the pinch-off charge density are both affected by small-geometry effects. According to the description of small-geometry effects, the prime factor responsible for deviations in the threshold voltage is the variation of the depletion charge, which can be associated with an equivalent variation in the body effect factor. γ affects both the pinch-off voltage and the slope factor. The shift in γ in relation to a long and wide channel device can be written in terms of the threshold voltage variation as

$$\Delta\gamma \cong \frac{\Delta V_T}{\sqrt{2}\phi_F}. \quad (4.8.2)$$

Once the value of $\Delta\gamma$ is known, the values of the pinch-off voltage and slope factor can be modified accordingly and UCCM can be used to determine the inversion charge density.

In the circuit design-oriented approach to modeling the saturation region, the output conductance is assumed to be proportional to the drain current and inversely proportional to the Early voltage V_A [6], a constant parameter in first order models such as SPICE1. However, a constant Early voltage is inadequate to model the output conductance for the simulation of analog circuits.

A physics-based model of the output conductance includes velocity saturation effects, CLM, and DIBL. Impact ionization and substrate current-induced body effects can be easily included in the model by adding a substrate current, as in [43].

From (4.3.1) one can write:

$$g_o = \frac{dI_{Dsat}}{dV_D} = -Wv_{sat} \frac{dQ'_{IDsat}}{dV_D}, \quad (4.8.3)$$

where g_o is the output conductance in saturation, previously written as g_{md} . According to (4.3.4), the saturation charge Q'_{IDsat} depends on the modulated channel length, $L_e = L - \Delta L$, through ζ and on the inversion

charge density at source, Q'_{IS} . At this point we can include both the CLM and DIBL effects to calculate the small-signal output conductance.

Now, (4.8.3), (4.3.2), and (4.3.4), together with UCCM, which gives us the relationship between voltage and charge, allow the calculation of the MOSFET output conductance-to-current ratio:

$$\frac{g_o}{I_{Dsat}} = \frac{1}{V_A} = \frac{1}{V_{ADIBL}} + \frac{1}{V_{ACLM}}. \quad (4.8.4)$$

The two components of the Early voltage can be calculated from

$$\frac{I_{Dsat}}{V_{ADIBL}} = \left. \frac{dI_{Dsat}}{dV_D} \right|_{\Delta L=0} = \frac{dI_{Dsat}}{dq'_{IS}} \frac{dq'_{IS}}{dV_P} \frac{dV_P}{dV_T} \frac{dV_T}{dV_D} \quad (4.8.5)$$

and

$$\frac{I_{Dsat}}{V_{ACLM}} = \left. \frac{dI_{Dsat}}{dV_D} \right|_{\sigma=0} = \frac{dI_{Dsat}}{d\Delta L} \frac{d\Delta L}{dV_D}. \quad (4.8.6)$$

The first term in the right-hand side of (4.8.5) is calculated from (4.3.5), and the second from UCCM, which is repeated here in its normalized form

$$V_P - V_{S(D)} = \phi_t \left[q'_{IS(D)} - 1 + \ln q'_{IS(D)} \right]. \quad (4.8.7)$$

The third term in (4.8.5) is equal to $-1/n$ while the fourth is, from (4.7.6), equal to $-\sigma$. The Early voltage associated with DIBL results in

$$\frac{V_{ADIBL}}{\phi_t} \cong \frac{n}{2\sigma} \left(\sqrt{1 + i_{Dsat}} + 1 \right). \quad (4.8.8)$$

The CLM Early voltage is calculated from (4.4.2) and (4.4.1), yielding

$$\frac{V_{ACLM}}{\phi_t} \cong \frac{1}{\phi_t} \left(\frac{1}{L_e} \frac{d\Delta L}{dV_D} \right)^{-1} = \frac{V_E}{\phi_t} \frac{L_e}{L_C} \left(1 + \frac{V_{DS} - V_{DSsat}}{V_E} \right). \quad (4.8.9)$$

for $V_{DS} > V_{DSsat}$, with $L_e = L - \Delta L$. Expressions (4.8.8) and (4.8.9) are valid for $i_{Dsat} < 1/\xi^2$.

The set of equations (4.8.3) through (4.8.9) is a generalization, for any bias condition, of the MOSFET output conductance presented in [44]. In weak inversion $i_{Dsat} < 1$; therefore, the Early voltage V_A is independent of the current level. If $1 \ll i_{Dsat} < 1/\zeta^2$, the DIBL component of the Early voltage is proportional to the square root of the current while the CLM component does not depend on the current level. Moreover, the CLM component of V_A depends on the effective voltage drop across the shrunk part of the channel. Typically, the DIBL component of the Early voltage can be neglected for high inversion levels, CLM being the dominant factor in the output conductance.

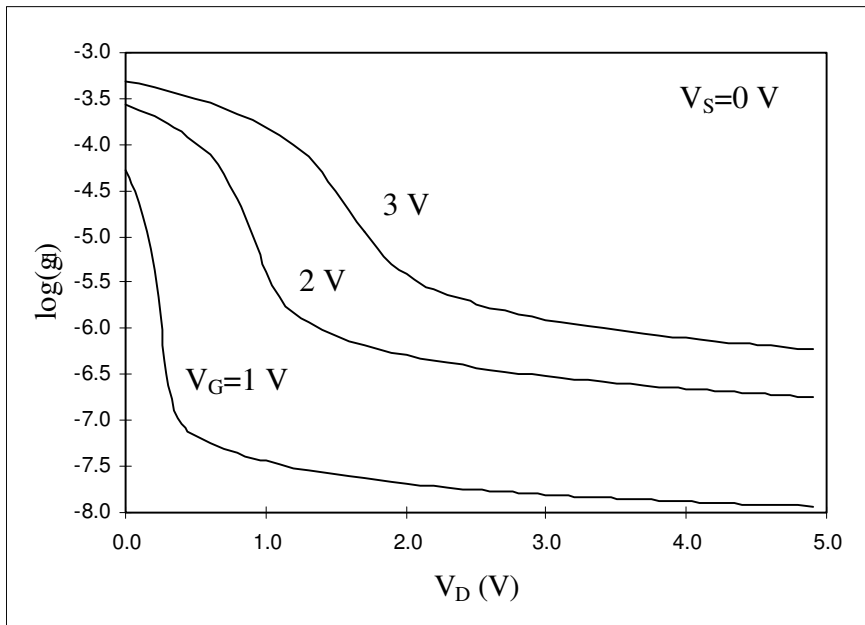


Fig. 4.12 MOSFET output conductance, simulated according to the model described by expressions (4.8.4), (4.8.8), and (4.8.9).

Fig. 4.12 illustrates the output conductance of a MOSFET, simulated according to the model described by equations (4.8.4), (4.8.8), and (4.8.9). A smoothing function has been used to avoid a discontinuous derivative of g_o with respect to the voltage in the transition from non-saturation to saturation.

In order to verify the consistence of the output conductance model, MOSFETs from a $0.75\mu\text{m}$ technology with different channel lengths were measured for several bias conditions. Fig. 4.13 shows the variation in the Early voltage for several channel lengths in terms of the normalized drain current. One can conclude that the Early voltage increases with increasing channel length, is almost independent of the current level in weak inversion, and increases in moderate and strong inversion, as predicted by (4.8.8) and (4.8.9).

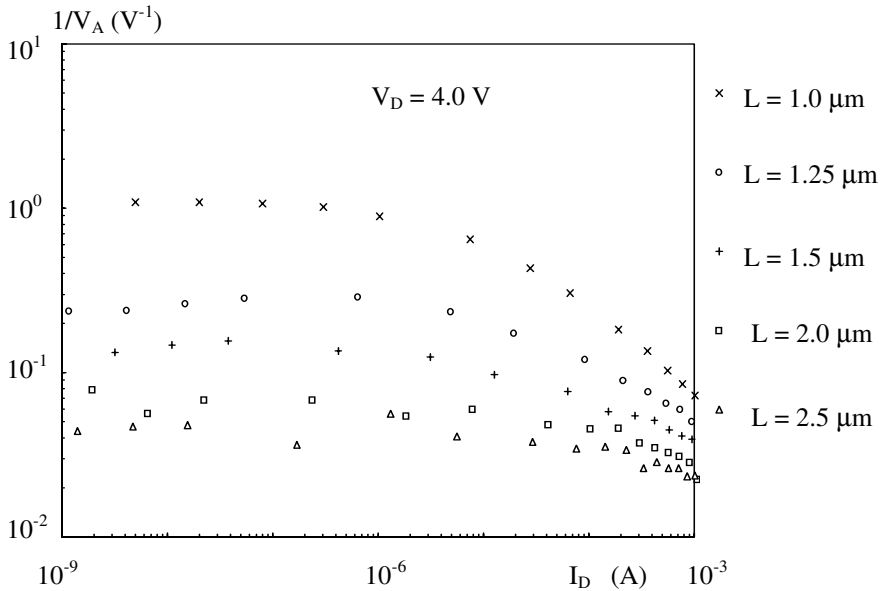


Fig. 4.13 Reciprocal of the Early voltage vs. drain current for MOSFETs in a $0.75 \mu\text{m}$ CMOS technology.

4.9 Benchmark tests for dc MOSFET models for circuit simulation

For many years, the modeling of MOS transistors for IC design has been driven by the needs of digital circuit simulation. However, the trend started in the late 70's toward mixed analog-digital design generated the necessity for MOSFET models appropriate for analog design as well. The more recent developments in RF CMOS integrated circuits have put more pressure on the need for improvement of the MOSFET model.

A good transistor model must “pass” a set of qualitative and quantitative tests [6], [45], [46]. The qualitative tests are used to determine the model capabilities and to show any shortcomings in the model equations [45], *e.g.*, transistor asymmetric behavior when source and drain are interchanged, or discontinuous derivatives. Quantitative tests evaluate how well the model represents the characteristics of real devices. The model should also have a physics base to allow for statistical analysis and predict matching behavior. The model equations must be simplified in such a way that users can understand them and to facilitate the procedure of parameter extraction. Finally, as pointed out in [1], a subset of the core model equations should be available for hand calculations. A simplified MOS model for hand calculations is extremely important to giving designers an intuitive understanding of circuits as well as providing a simple means for first order calculations of electrical variables.

Some of the examples of qualitative testing of compact models shown in [47] will be given next. These examples were run in SMASH [48] to show dc, ac, and transient behavior of the Advanced Compact MOSFET (ACM) model and some popular MOSFET models used in the 1990s. Tests were performed for devices in a 0.8 μm CMOS technology.

4.9.1 Series-parallel association of transistors

In integrated circuits, it is very common to associate transistors in series and/or in parallel. Fig. 4.14 shows a series-parallel association of 16 identical transistors, all of them sharing a common substrate. Fig. 4.15 displays the corresponding dc output characteristics of this arrangement for two values of gate-to-source voltages. Also shown are the output characteristics of a single transistor which is identical to any of the individual transistors in the arrangement. For low drain-to-source voltages the transistor behaves as a linear resistor; therefore, the series-parallel association of linear resistors also holds for transistors. Note that, in accordance with the previous argument, the characteristics of an individual transistor and of the arrangement are coincident for low values of drain-to-source voltage. For higher drain voltages, the current in the single transistor is more sensitive than that of the arrangement to

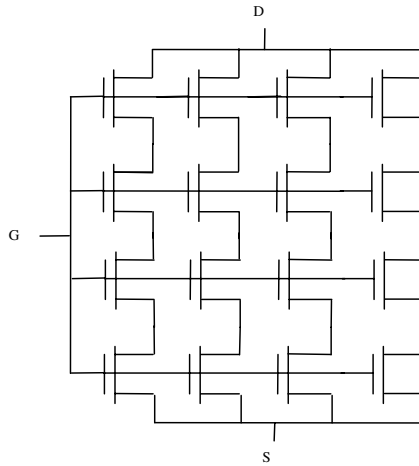


Fig. 4.14 Series-parallel association of transistors in a 4 by 4 arrangement. All transistors are identical ($W=10\text{ }\mu\text{m}$ and $L=2\text{ }\mu\text{m}$).

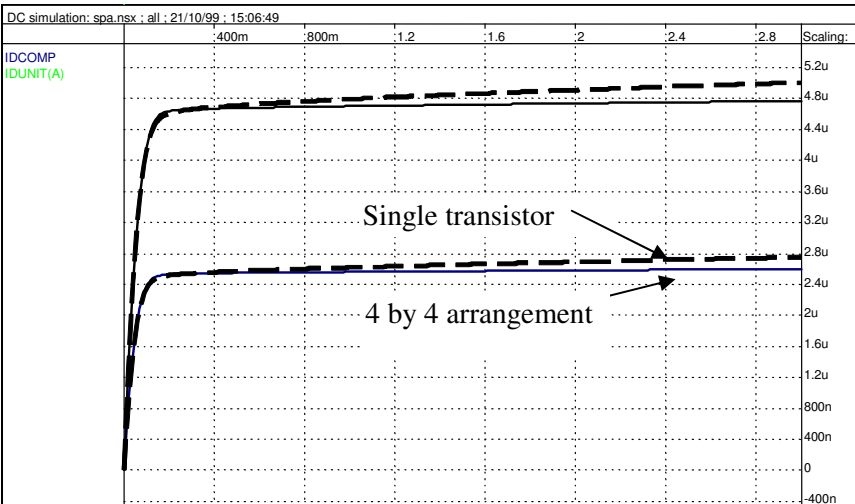


Fig. 4.15. DC output characteristics of the 4 by 4 transistor array in Fig. 4.14 and of a single transistor using the ACM model.

variations in the drain voltage. The reason for this discrepancy is not a model failure but short-channel effects.

MOSFET current dividers are useful components of analog circuits that can also be used to test the quality of compact models. Fig. 4.16 shows one such divider. In this array, all transistors have the same dimensions and share a common substrate. A first order analysis of this topology shows that the reference current is successively divided by two.

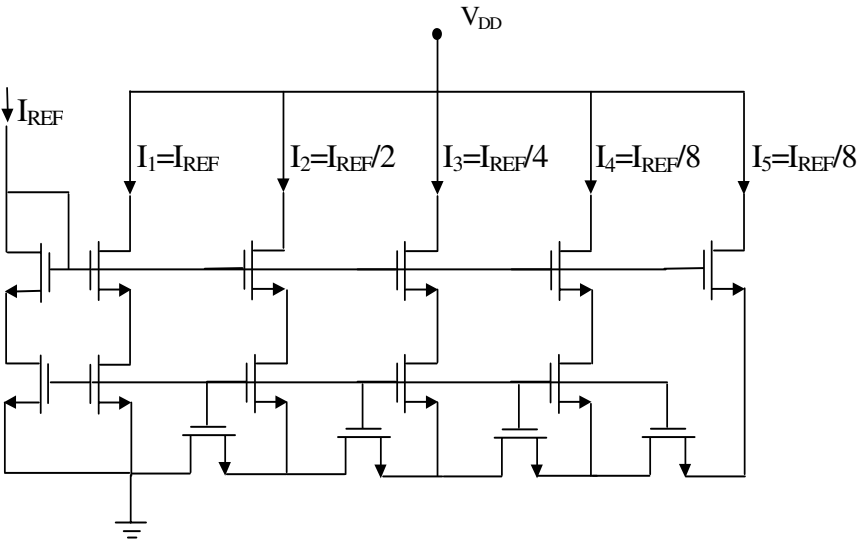


Fig. 4.16 MOSFET binary current divider ($W=100\text{ }\mu\text{m}$, $L=20\text{ }\mu\text{m}$).

In order to reduce deviations from the expected values owing to short-channel effects, long-channel devices have been employed in the current divider. Fig. 4.17 shows that the maximum normalized error using the ACM model does not exceed 0.15% for a 3-decade variation of the reference current I_{REF} . The normalized currents show a consistent reduction with increasing reference currents; this phenomenon is explained by noting that the voltage at the input node (at the leftmost branch) gets closer to V_{DD} for increasing reference currents. Fig. 4.18 presents simulation results for the same current divider using the BSIM3v3 MOS transistor model; in this case, the normalized errors can

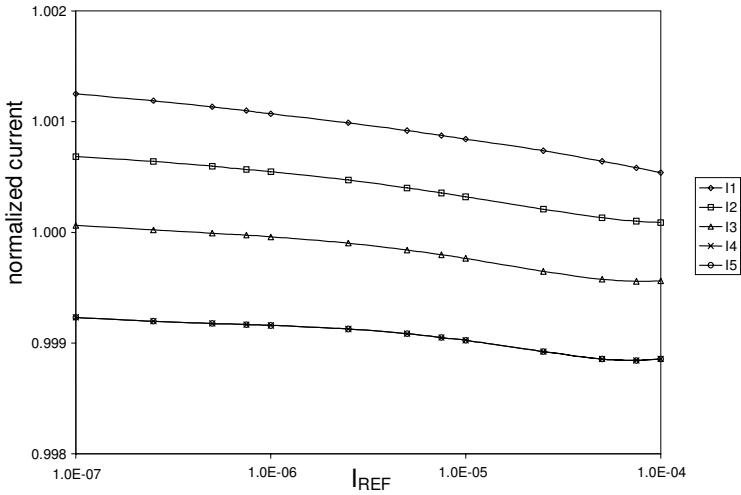


Fig. 4.17 Normalized branch currents vs. input reference current obtained from simulations using the ACM model in SMASH.

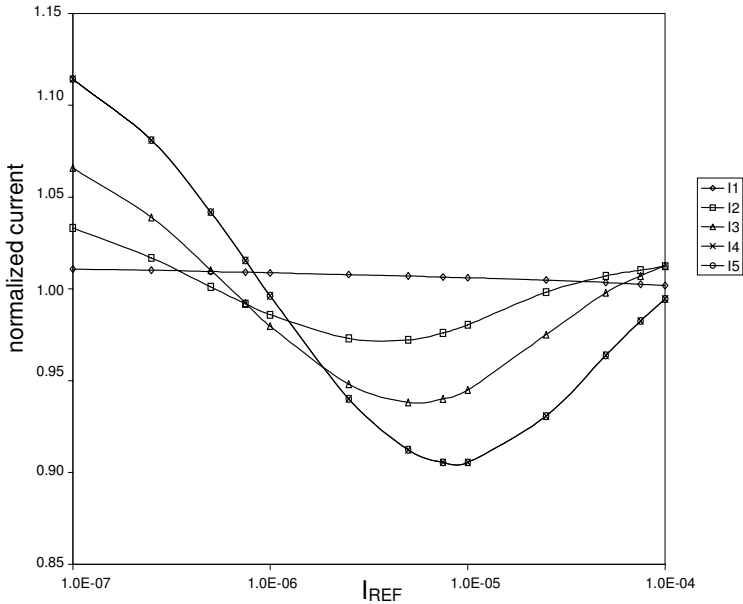


Fig. 4.18 Normalized branch currents vs. input reference current obtained from simulations using the BSIM 3v3 model in SMASH.

be as high as 10%. Not only are the errors exaggerated but also their oscillatory behavior, which can be attributed to poor modeling of the MOS transistor.

4.9.2 Gummel symmetry test

Integrated MOS transistors are symmetric devices; source and drain can be interchanged but the transistor characteristics should remain the same. The circuit for the Gummel symmetry test is shown in Fig. 4.19 [45].

Since the MOSFET is a symmetric device, then $I_D(V_X) = -I_D(-V_X)$, which requires the even order derivatives to be odd functions of V_X and, as a result, the second order derivative to be zero around $V_{DS}=0$. Gummel tests using the ACM and BSIM3v3 are shown in Fig. 4.20 and Fig. 4.21, respectively. The ACM model passes the Gummel symmetry test, while BSIM3v3 fails the test because the device characteristic is not symmetric around $V_{DS}=0$.

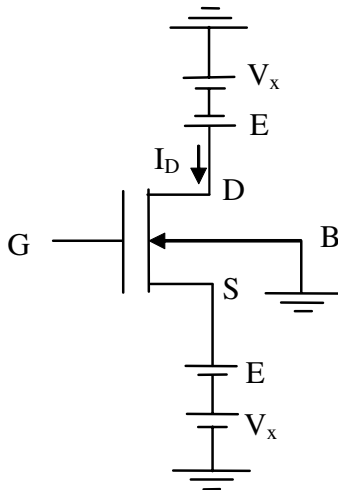


Fig. 4.19 Circuit for the Gummel symmetry test.

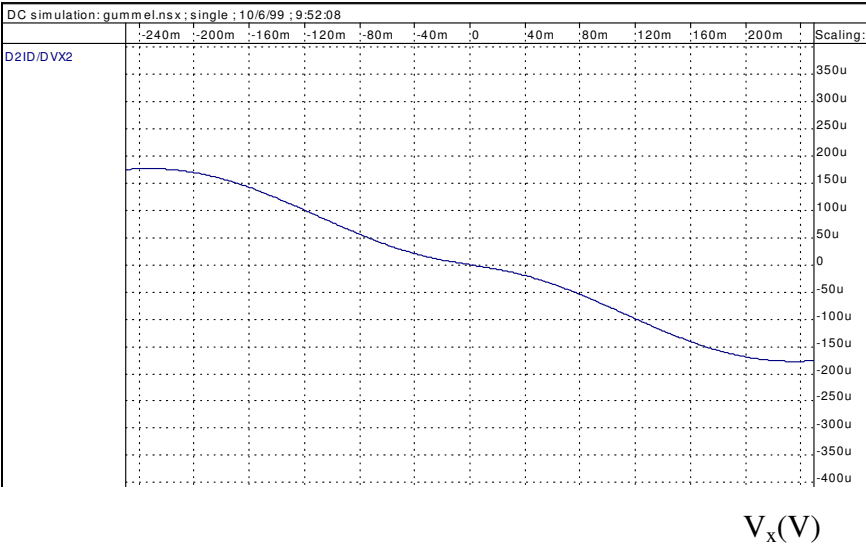


Fig. 4.20 Second order derivative of the drain current with respect to V_x using the ACM model.

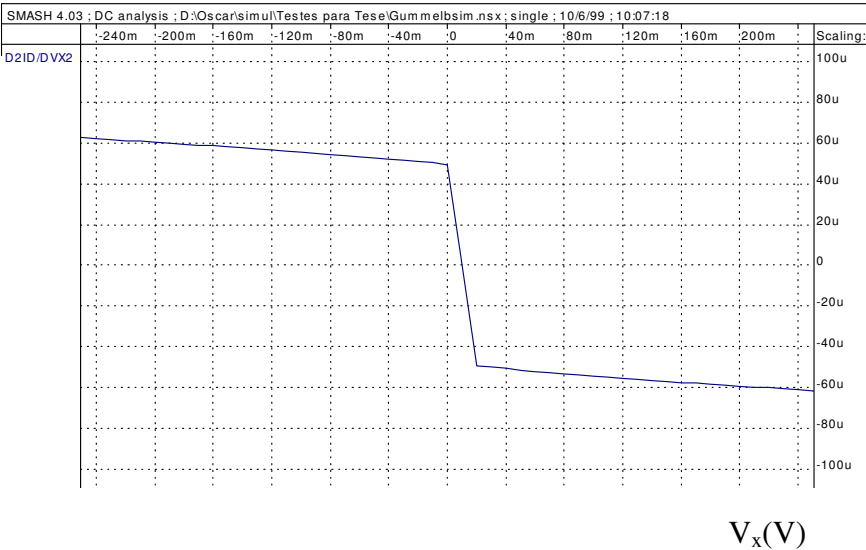


Fig. 4.21 Second order derivative of the drain current with respect to V_x using the BSIM3v3 model.

Bibliography

- [1] Daniel P. Foty, *MOSFET Modeling with SPICE - Principles and Practice*, Prentice Hall, Upper Saddle River, NJ, 1997
- [2] J. R. Schrieffer, "Effective carrier mobility in surface-space charge layers," *Phys. Rev.*, vol. 97, no. 3, pp. 641-646, Feb. 1955.
- [3] S. C. Sun and James D. Plummer, "Electron mobility in inversion and accumulation layers on thermally oxidized silicon surfaces," *IEEE Trans. Electron Devices*, vol. 27, no. 8, pp. 1497-1508, Aug. 1980.
- [4] Shin-Ichi Takagi, Akira Toriumi, Masao Iwase, and Hiroyuki Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I - Effects of substrate impurity concentration," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2357-2362, Dec. 1994.
- [5] Mark Lundstrom, *Fundamentals of Carrier Transport*, 2nd ed., Cambridge University Press, Cambridge, UK, 2000.
- [6] Yannis Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed., McGraw-Hill, Boston, 1999.
- [7] Robert F. Pierret, *Advanced Semiconductor Fundamentals – Vol. VI of Modular Series on Solid State Devices*, Addison-Wesley, Reading, Mass., 1987.
- [8] Anant G. Sabnis and James T. Clemens, "Characterization of electron mobility in the inverted <100> Si surface," *IEDM Tech. Dig.*, 1979, pp. 18-21.
- [9] Shin-Ichi Takagi, Akira Toriumi, Masao Iwase, and Hiroyuki Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part II - Effects of surface orientation," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2363-2368, Dec. 1994.
- [10] Matthias Bucher, *Analytical MOS Transistor Modelling for Analog Circuit Simulation*, PhD Thesis, EPFL, Lausanne, Switzerland, 1999.
- [11] Oscar da Costa Gouveia Filho, Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup Montoro, "Advanced compact model for short-channel MOS transistors," *Proc. CICC*, 2000, pp. 209-212.
- [12] Steven A. Schwarz and Stephen E. Russek, "Semi-empirical equations for electron velocity in silicon: Part II – MOS inversion layer," *IEEE Trans. Electron Devices*, vol. 30, no. 12, pp. 1634-1639, Dec. 1983.
- [13] Narain Arora, *MOSFET Models for VLSI Circuit Simulation Theory and Practice*, Springer-Verlag, Wien, 1990.
- [14] Benjamin Iñíguez and Eugenio García Moreno, "An improved C_{∞} continuous small-geometry MOSFET modeling for analog applications," *J. Analog Integr. Circ. Signal Process.*, vol. 13, no. 3, pp. 241-259, July 1997.
- [15] C. Canali, G. Majni, R. Minder, and G. Ottaviani, "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature," *IEEE Trans. Electron Devices*, vol. 22, no. 11, pp. 1045-1047, Nov. 1975.
- [16] R. W. Coen and R. S. Muller, "Velocity of surface carriers in inversion layers on silicon," *Solid-State Electron.*, vol. 23, no. 1, pp. 35-40, Jan. 1980.

- [17] Kwyro Lee, Michael Shur, Tor A. Fjeldly, and Trond Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [18] Mary Ann Maher, *A charge-controlled model for MOS transistors*, PhD Thesis, Caltech, Pasadena, CA, 1989.
- [19] Mary Ann Maher and Carver A. Mead, "A physical charge-controlled model for MOS transistors," in *Advanced Research in VLSI*, P. Losleben (ed.), MIT Press, Cambridge, MA, 1987.
- [20] M. C. Schneider, C. Galup-Montoro, O. C. Gouveia Filho, and A. I. A. Cunha, "A single-piece charge-based model for the output conductance of MOS transistors," *Proc. IEEE Int. Conf. Electron. Circ. Syst.*, 1998, pp. 545-548.
- [21] Y. A. El-Mansy and A. R. Boothroyd, "A simple two-dimensional model for IGFET operation in the saturation region," *IEEE Trans. Electron Devices*, vol. 24, no. 3, pp. 254-262, Mar. 1977.
- [22] P. K. Ko, R. S. Muller, and C. Hu, "A unified model for hot-electron current in MOSFETs," *IEDM Tech. Dig.*, 1981, pp. 600-603.
- [23] Kenji Natori, "Ballistic metal-oxide-semiconductor field effect transistor," *J. Appl. Phys.*, vol. 76, no.8, pp. 4879-4890, 15 Oct. 1994.
- [24] Mark Lundstrom, "Elementary scattering theory of the Si MOSFET," *IEEE Electron Device Lett.*, vol. 18, no.7, pp. 361-363, July 1997.
- [25] Mark Lundstrom and Zhibin Ren, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 133-141, Jan. 2002.
- [26] Chau-Ren Ji and Chih-Tang Sah, "Two-dimensional numerical analysis of the narrow gate effect in MOSFET," *IEEE Trans. Electron Devices*, vol. 30, no. 6, pp. 635-647, June 1983.
- [27] L. D. Yau, "A simple theory to predict the threshold voltage of short-channel IGFETs," *Solid-State Electron.*, vol. 17, no. 10, pp. 1059-1063, Oct. 1974.
- [28] L. A. Akers and C. S. Chao, "A closed-form threshold voltage expression for a small-geometry MOSFET," *IEEE Trans. Electron Devices*, vol. 29, no. 4, pp. 776-778, Apr. 1982.
- [29] Yuan Taur, Genda J. Hu, Robert H. Dennard, Lewis M. Terman, Chung-Yu Ting, and Karen E. Petrillo, "A self-aligned 1- μ m-channel CMOS technology with retrograde n-well and thin epitaxy," *IEEE Trans. Electron Devices*, vol. 32, no. 2, pp. 203-209, Feb. 1985.
- [30] K. N. Ratnakumar and James D. Meindl, "Short-channel MOST threshold voltage model," *IEEE J. Solid-State Circuits*, vol. 17, no. 5, pp. 937-948, Oct. 1982.
- [31] Rafael Rios, Wei-Kai Shih, Atul Shah, Sivakumar Mudanai, Paul Packan, Tracy Sandford, and Kaizad Mistry, "A three-transistor threshold voltage model for halo processes," *IEDM Tech. Dig.*, 2002, pp. 113-116.
- [32] Narain D. Arora and Mahesh S. Sharma, "Modeling the anomalous threshold voltage behavior of submicron MOSFETs," *IEEE Electron Device Lett.*, vol.13, no.2, pp. 92-94, Feb. 1992.
- [33] Christian C. Enz, François Krummenacher and Eric A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *J. Analog Integr. Circuits Signal Process.*, vol. 8, pp. 83-114, July 1995.

- [34] Lex A. Akers, Moiz M. E. Beguwalla, and Frank Z. Custode, "A model of a narrow-width MOSFET including tapered oxide and doping encroachment," *IEEE Trans. Electron Devices*, vol. 28, no. 12, pp. 1490-1495, Dec. 1981.
- [35] Ronald R. Troutman, "VLSI limitations from drain-induced barrier lowering," *IEEE Trans. Electron Devices*, vol. 26, no. 4, pp. 461-469, Apr. 1979.
- [36] R. R. Troutman and A. G. Fortino, "Simple model for threshold voltage in a short-channel IGFET," *IEEE Trans. Electron Devices*, vol. 24, no. 10, pp. 1266-1268, Oct. 1977.
- [37] Savvas G. Chamberlain and Sannasi Ramanan, "Drain-induced barrier-lowering analysis in VLSI MOSFET using two-dimensional numerical simulations," *IEEE Trans. Electron Devices*, vol. 33, no. 11, pp. 1745-1753, Nov. 1986.
- [38] C. R. Viswanathan, Bruce C. Burke, Gerrit Lubberts, and Timothy J. Tredwell, "Threshold voltage in short-channel MOS devices," *IEEE Trans. Electron Devices*, vol. 32, no. 5, pp. 932-940, May 1985.
- [39] Joseph M. Pimbley and James D. Meindl, "MOSFET scaling limits determined by subthreshold conduction," *IEEE Trans. Electron Devices*, vol. 36, no. 9, pp. 1711-1721, Sep. 1989.
- [40] Zhi-Hong Liu, Chenming Hu, Jian-Hui Huang, Tung-Yi Chan, Min-Chie Jeng, Ping K. Ko, and Y. C. Cheng, "Threshold voltage for deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 40, no. 1, pp. 86-95, Jan. 1993.
- [41] Jiin-Jang Maa and Ching-Yuan Wu, "A new simplified threshold-voltage model for n-MOSFET's with nonuniformly doped substrate and its application to MOSFET's miniaturization," *IEEE Trans. Electron Devices*, vol. 42, no. 8, pp. 1487-1494, Aug. 1995.
- [42] Yuan Taur and Tak H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.
- [43] Yuhua Cheng, Min-Chie Jeng, Zhihong Liu, Jianhui Huang, Mansun Chan, Kai Chen, Ping Keung Ko, and Chenming Hu, "A physical and scalable I-V model in BSIM3v3 for analog/digital simulation," *IEEE Trans. Electron Devices*, vol. 44, no. 2, pp. 277-287, Feb. 1997.
- [44] Kai-Yap Toh, Ping-Keung Ko, and Robert G. Meyer, "An engineering model for short-channel MOS devices," *IEEE J. Solid-State Circuits*, vol. 23, no. 4, pp. 950-958, Aug. 1988.
- [45] IEEE Recommended Practices #P1485 on: Test Procedures for Micro-electronic MOSFET Circuit Simulator Model Validation, 1997. [Online] Available: <http://ray.eeel.nist.gov/modval>
- [46] Yannis P. Tsividis and Ken Suyama, "MOSFET modeling for analog circuit CAD: problems and prospects," *IEEE J. Solid-State Circuits*, vol. 29, no. 3, pp. 210-216, Mar. 1994.
- [47] Oscar da Costa Gouveia Filho, *Um modelo compacto do transistor MOS para simulação de circuitos*, PhD Thesis, Federal University of Santa Catarina, Florianopolis, Brazil, 1999 (in Portuguese).
- [48] SMASH Users Manual, Dolphin Integration, Meylan, France, 1995.

Problems

4.1. To solve this problem, assume the following data: $\mu_s = 250 \text{ cm}^2/\text{Vs}$, $C'_{ox} = 16 \text{ fF}/\mu\text{m}^2$, $\phi_t = 25 \text{ mV}$, $F_C = 40 \text{ kV/cm}$ and two transistors whose aspect ratios are $W_1/L_1 = 10 \mu\text{m}/10 \mu\text{m}$ and $W_2/L_2 = 0.10 \mu\text{m}/0.10 \mu\text{m}$.

a) Integrate equation (4.2.5) to find (4.2.6).

b) Assuming that both transistors are characterized by the I - Q equation in (4.2.6), plot the ratio I_{D1}/I_{D2} for $0 \leq Q'_{ID} - Q'_{IS} \leq 2 \cdot 10^{-2} \text{ C/m}^2$.

c) For a saturation current of the short-channel device equal to $15.6 \mu\text{A}$, find the corresponding value of the inversion charge density at source Q'_{IS} and the (saturated) inversion charge density at drain Q'_{ID} . Assuming $V_S = 0$, determine the pinch-off voltage V_P and the drain-to-source saturation voltage.

4.2. a) Neglecting the variation of n and μ_s with V_G , show that

$$g_{mg} \cong \mu_s C'_{ox} \phi_t \frac{W}{L} \frac{\sqrt{1+i_{Dsat}}}{1+\zeta\sqrt{1+i_{Dsat}}} \frac{(\sqrt{1+i_{Dsat}} - 1 + \zeta i_{Dsat}/2)}{\sqrt{1+i_{Dsat}} + \zeta i_{Dsat}/2}. \quad (\text{P4.2.1})$$

b) Using approximations $\zeta \ll 1$ and $1 + \zeta\sqrt{1+i_{Dsat}} \cong 1 + \zeta\sqrt{i_{Dsat}}$, deduce equation (4.3.6).

c) For very high currents ($i_{Dsat} > 1/\zeta^2$), demonstrate that (4.3.6) tends asymptotically toward (4.3.7).

4.3. Consider the drain-to-source saturation voltage written as in (4.3.8). Plot V_{DSsat} vs. $\log(Q'_{IS}/-nC'_{ox}\phi_t)$ for $0.01 \leq (Q'_{IS}/-nC'_{ox}\phi_t) \leq 100$ and $\zeta = 0.001, 0.01$, and 0.1 .

4.4. Derive the equations for the Early voltages associated with the DIBL and CLM effects. Assuming that $V_{DS} \ll V_E$, calculate the normalized saturation current such that the Early voltages due to CLM and DIBL are equal. Propose a method on how to measure the values of V_{ADIBL} and V_{ACLM} assuming that expressions (4.8.8) and (4.8.9) are valid and that V_{ACLM} is almost insensitive to V_{DS} .

4.5. Assuming that for constant drain voltage the threshold voltage variation due to DIBL is modeled as $V_T = V_{T0} - \sigma V_S$ where σ is the DIBL factor, show that, in weak inversion saturation, $g_{ms} \phi_t / I_D = 1 - \sigma / n$.

This page intentionally left blank

Chapter 5

Stored Charges and Capacitive Coefficients

The drain current expressions of the previous chapter were deduced assuming steady-state (dc) operation. In most circuits the terminal voltages of the device, of course, vary over time. In dynamic operation, the charges stored in the device vary and (charging) currents flow into the device to build up the varying stored charges.

In this chapter we will calculate the stored charges and the terminal current under large-signal dynamic operation. We will analyze the part of the device between the source and drain enclosed in the broken-line rectangle of Fig. 5.1. This part is called the intrinsic transistor and is where the field effect occurs.

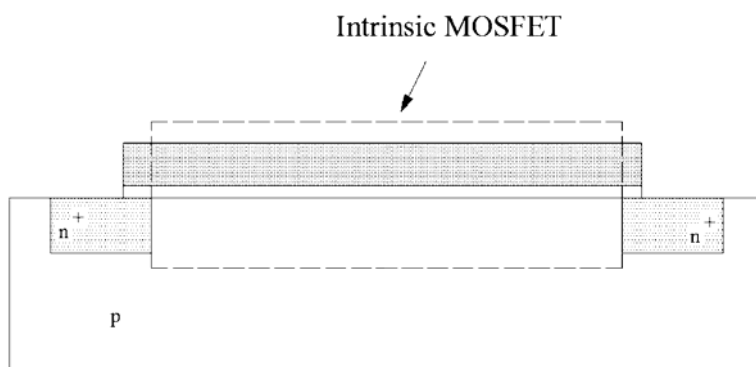


Fig. 5.1 MOS transistor structure showing the intrinsic part.

The rest of the device, which includes the source- and drain-substrate junctions, and the overlap between the gate and the source and drain regions, is the extrinsic part of the transistor. The main objective of

transistor modeling is to develop models of the intrinsic transistor. The extrinsic element (*e.g.*, junctions, capacitances) models are added afterwards to represent the effects of the extrinsic parts on transistor performance.

5.1 Transient analysis of the MOS transistor [1]

The development of accurate dynamic models of the MOS transistors has taken much longer than that of the bipolar transistor. For a while the distributed gate-channel capacitances were modeled by three lumped capacitances (C_{gs} , C_{gd} , C_{gb}) as proposed by Meyer [2]. Since the MOSFET is a four-terminal device, the three-capacitor model is inappropriate because for a four-terminal device nine independent coefficients are necessary to model the relationship between three independent voltages and currents.

Consistent dynamic MOSFET models are based on the pioneering work of Oh, Ward, and Dutton [1] of the late 70's that will now be summarized.

5.1.1 The continuity equation [6]

The continuity equation [3] (Appendix D) is readily derived for the MOSFET, considering laminar flow and neglecting generation and recombination. The charge increase in a piece of the inversion layer of length Δy in an interval Δt is the difference between the charge entering the right side and the charge leaving the left side (see Fig. 5.2). The charge increase can also be calculated as the charge density increase times the area of the piece of inversion layer. Thus,

$$(I + \Delta I)\Delta t - I\Delta t = \Delta I \cdot \Delta t = \Delta Q'_I (W \Delta y) \quad (5.1.1)$$

or

$$W \frac{\Delta Q'_I}{\Delta t} = \frac{\Delta I}{\Delta y} . \quad (5.1.2)$$

Considering the limit case $\Delta y \rightarrow 0$ and interval $\Delta t \rightarrow 0$, the continuity equation follows

$$W \frac{\partial Q'_I(y,t)}{\partial t} = \frac{\partial I(y,t)}{\partial y} \quad (5.1.3)$$

where $I(y,t)$ is the time-varying channel current, no longer assumed to be constant along the channel length, and $Q'_I(y,t)$ is the time-varying inversion charge density. An accurate physics-based model of the dynamic behavior of the MOSFET requires the resolution of the above partial differential equation that represents the distributed nature of the MOSFET. The problem is analytically involved and will be considered in the next section. In the rest of this chapter, approximate solutions of the continuity equation [1] will be employed to develop compact dynamic MOSFET models.

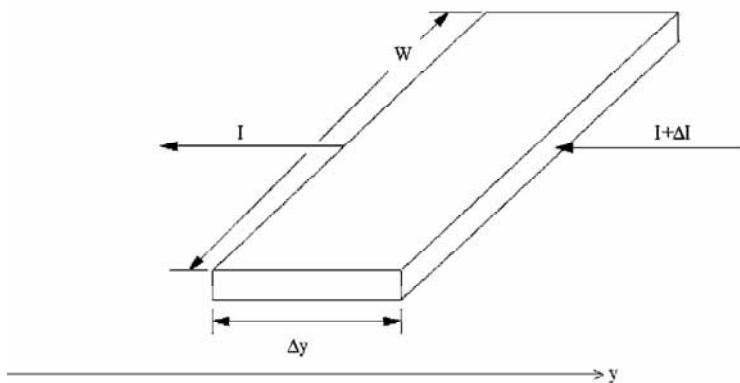


Fig. 5.2 Section of the inversion layer of width W and length Δy .

5.1.2 Definitions of source and drain charges [1]

This section presents the derivation of simple expressions for the transient drain and source currents in terms of the transport current in the channel and the stored charge in the device. In order to obtain consistent ac and dc models, the same basic approximations used in the derivation of the dc current expressions are adopted: gradual channel approximation and laminar (parallel to the surface) current. Thus, the time-varying current transport expression, analogous to the dc expression (3.3.9), may be written as

$$I(y,t) = -\mu_n W Q'_I(y,t) \frac{\partial V_C(y,t)}{\partial y} \quad (5.1.4)$$

where $I(y,t)$ is the time-varying channel current, no longer assumed to be constant along the channel, flowing in the minus y direction (from drain to source). It is important to remark that (5.1.4) considers the distributed structure of the transistor, with the current, the inversion charge density and the channel potential being functions of both position in the channel y and time t . The inversion channel charge density is obtained from the unidimensional Poisson's equation, assuming that the gradual channel approximation is valid and that the device is uniform in cross-section between the source and the drain. This implies that the inversion charge density may be written as

$$Q'_I(y,t) = Q'_I(V_C(y,t), V_G(t)) = Q'_I(V_C, V_G). \quad (5.1.5)$$

The value of Q'_I is dependent on space and time only through its dependence on both the channel potential $V_C(y, t)$ and the gate voltage $V_G(t)$ as long as the gradual channel approximation is valid, the device is uniform in cross section, and the time constants associated with fast surface states and dielectric relaxation can be ignored [1]. Expression (5.1.5) is equivalent to saying that the inversion charge density at each point of the channel varies synchronously with the gate voltage and with the channel voltage at that point of the channel.

To calculate the transient source current $I_S(t)=I(0,t)$ we can proceed as follows [1]. Equation (5.1.3) is integrated from the source ($y=0$) to position y along the channel giving $I(y,t)$. Writing the current $I(y,t)$ using (5.1.4) yields

$$W \int_0^y \frac{\partial Q'_I(\hat{y},t)}{\partial t} d\hat{y} = -\mu_n W Q'_I(y,t) \frac{\partial V_C(y,t)}{\partial y} - I(0,t). \quad (5.1.6)$$

Integrating again from source ($y=0$) to drain ($y=L$) and solving for the source current $I(0,t)$ gives

$$I_S(t) = I(0, t) = -\mu_n \frac{W}{L} \int_{V_S(t)}^{V_D(t)} Q'_I(V_C, V_G) dV_C - \frac{W}{L} \int_0^L \int_0^y \frac{\partial Q'_I(\hat{y}, t)}{\partial t} d\hat{y} dy. \quad (5.1.7)$$

The first integral in the right-hand side of (5.1.7) results from changing the integration variable from length to potential (as allowed by (5.1.5)). The transport current is given by

$$I_T(t) = -\mu_n \frac{W}{L} \int_{V_S(t)}^{V_D(t)} Q'_I(V_C) dV_C. \quad (5.1.8)$$

Consequently, we can rewrite (5.1.7) as

$$I_S(t) = I_T(t) - \frac{dQ_S}{dt}. \quad (5.1.9)$$

Thus, the transient source current can be written as the sum of a transport current, which reflects the time dependence of the boundary conditions only [1], and a displacement (source charging) current, for which the source charge is defined as

$$Q_S = \frac{W}{L} \int_0^L \int_0^y Q'_I d\hat{y} dy. \quad (5.1.10)$$

Using integration by parts as indicated below

$$\int_0^L \int_0^y Q'_I d\hat{y} dy = y \int_0^y Q'_I d\hat{y} \Big|_0^L - \int_0^L y Q'_I dy = \int_0^L (L-y) Q'_I dy, \quad (5.1.11)$$

the source stored charge Q_S can be written as

$$Q_S = W \int_0^L \left(1 - \frac{y}{L}\right) Q'_I dy. \quad (5.1.12)$$

The drain current is found integrating (5.1.3) from source ($y=0$) to drain ($y=L$) and calculating $I(0, t) = I_S(t)$ in the resulting equation using (5.1.9) and (5.1.12). Thus, we obtain

$$I_D(t) = I_T(t) + \frac{dQ_D}{dt} \quad (5.1.13)$$

where

$$Q_D = W \int_0^L \frac{y}{L} Q'_I dy. \quad (5.1.14)$$

As expected,

$$I_D(t) - I_S(t) = \frac{dQ_D}{dt} + \frac{dQ_S}{dt} = \frac{dQ_I}{dt}, \quad (5.1.15)$$

where

$$Q_I = W \int_0^L Q'_I dy \quad (5.1.16)$$

is the total inversion charge stored in the channel. The difference between the drain and source currents is exactly the amount required to change the inversion charge in the channel. The symmetry of the expressions for drain and source charges is self-evident.

Concerning the transient gate and bulk charges, the (leakage) transport current is usually neglected and in this case the transient currents reduce to the charging currents

$$I_G(t) = \frac{dQ_G}{dt} \quad (5.1.17)$$

and

$$I_B(t) = \frac{dQ_B}{dt}, \quad (5.1.18)$$

where

$$\begin{aligned} Q_G &= W \int_0^L Q'_G dy \\ Q_B &= W \int_0^L Q'_B dy. \end{aligned} \quad (5.1.19)$$

To obtain closed-form solutions for the transient current, the so-called quasi-static approximation can be used. In the quasi-static approximation it is assumed that the charge distribution in the channel is given by the dc distribution calculated considering the instantaneous values of the terminal voltages. Clearly this approximation fails for rapid changes in

the terminal voltages because the charge distribution in the channel cannot follow the terminal voltages instantaneously owing to the distributed nature of the channel.

We will derive in the next paragraphs the charge expressions for long-channel transistors and then include short-channel effects. All the information necessary to calculate the stored charges in the quasi-static approximation is available in the dc current model of the transistor.

5.2 Quasi-static charge-conserving model

In this section, single piece expressions for the stored charges, valid from weak, through moderate, to strong inversion are presented.

5.2.1 Stored charges [4]

The stored charges given by expressions, (5.1.12), (5.1.14), (5.1.16), and (5.1.19) are easily calculated changing the integration variable from length y to inversion charge density Q'_I . The two relationships needed to relate dy to dQ'_I are repeated below.

The charge-sheet current expression, equation (3.4.21), is

$$I_D = I_{drift} + I_{diff} = -\mu_n W Q'_I \frac{d\phi_s}{dy} + \mu_n W \phi_t \frac{dQ'_I}{dy} \quad (5.2.1)$$

while the approximate linear relationship between Q'_I and ϕ_s for a given V_G , equation. (3.4.22), is

$$dQ'_I = nC'_{ox} d\phi_s. \quad (5.2.2)$$

Substituting the basic approximation stated by (5.2.2) into expression (5.2.1) of the drain current, we have

$$dy = -\frac{\mu_n W}{nC'_{ox} I_D} (Q'_I - nC'_{ox} \phi_t) dQ'_I, \quad (5.2.3)$$

which can be readily used to calculate the integrals for the total charges. It is convenient to define the new variable

$$Q'_{It} = Q'_I - nC'_{ox} \phi_t \quad (5.2.4)$$

by way of which a very compact symbolic model of the MOSFET is achieved.

5.2.1.1 Symbolic MOSFET model

Since n is constant along the channel (it depends only on V_{GB}), then

$$dQ'_L = dQ'_I \quad (5.2.5)$$

and (5.2.3) may be rewritten as

$$dy = -\frac{\mu_n W}{nC'_{ox} I_D} Q'_L dQ'_L. \quad (5.2.6)$$

Equation (5.2.6) includes the drift and diffusion mechanisms, but in terms of the new virtual charge variable Q'^1_L , the relationship between virtual charge and length y (5.2.6) is formally the same as in the classical strong inversion (SI) model, which considers only the drift component of the current. Integrating (5.2.6) between source ($y=0$) and drain ($y=L$) yields

$$I_D = \frac{\mu_n W}{C'_{ox} L} \frac{Q'^2_F - Q'^2_R}{2n}, \quad (5.2.7)$$

where Q'_F and Q'_R are the values of Q'_L calculated at source and drain, respectively

$$Q'_{F(R)} = Q'_{IS(D)} - nC'_{ox} \phi_t. \quad (5.2.8)$$

Defining a (forward) saturation current I_{D0} and a saturation coefficient α , (5.2.7) may be rewritten as

$$I_D = I_{D0}(1 - \alpha^2) \quad (5.2.9)$$

where

$$I_{D0} = \frac{\mu_n W Q'^2_F}{C'_{ox} L(2n)} \quad (5.2.10)$$

and

¹ A more formal and general definition for the virtual charge density is given in Section 5.4.1.

$$\alpha = \frac{Q'_R}{Q'_F} = \frac{Q'_{ID} - nC'_{ox}\phi_t}{Q'_{IS} - nC'_{ox}\phi_t}. \quad (5.2.11)$$

The saturation coefficient α above is a generalization of the definition proposed in [5]. In effect, in SI and considering zero bulk charge, the relations given below hold.

$$\alpha \cong \frac{Q'_{ID}}{Q'_{IS}} \cong \frac{V_P - V_D}{V_P - V_S} \cong \frac{V_{GD} - V_{T0}}{V_{GS} - V_{T0}}. \quad (5.2.12)$$

The last term in (5.2.12) was used to define α in [5]. The advantage of the charge-based definition in (5.2.11) for the saturation coefficient is its generality.

Following the normal convention in which the source is more heavily inverted than the drain, $0 < \alpha \leq 1$. For $V_{DS}=0$, $\alpha=1$. In weak inversion, $\alpha \cong 1$ from the linear region to the saturation region. In strong inversion, α decreases from 1 at the origin ($V_{DS}=0$) to $\alpha \cong 0$ in saturation.

Summarizing, (5.2.9) is a single piece quadratic expression, formally the same as that for a transistor operating in the triode region in SI, but valid in weak, moderate and strong inversion, from the linear region to the saturation region.

5.2.1.2 Stored charge calculations

Substituting (5.2.4) and (5.2.6) into (5.1.16) yields:

$$Q_I = -\frac{\mu_n W^2}{I_D n C'_{ox}} \left[\int_{Q'_F}^{Q'_R} (Q'_I + nC'_{ox}\phi_t) Q'_I dQ'_I \right]. \quad (5.2.13)$$

The integration of (5.2.13) gives

$$Q_I = -\frac{\mu_n W^2}{nC'_{ox} I_D} \left[\frac{Q'^3_R - Q'^3_F}{3} + nC'_{ox}\phi_t \frac{Q'^2_R - Q'^2_F}{2} \right]. \quad (5.2.14)$$

Using expression (5.2.7) to substitute for the current I_D in the denominator of (5.2.14) and removing the common factor $(Q'_R - Q'_F)$ from both numerator and denominator, we obtain the total inversion channel charge as a rational function of the forward and reverse charge densities,

$$Q_I = WL \left[\frac{2}{3} \frac{Q_F'^2 + Q_F' Q_R' + Q_R'^2}{Q_F' + Q_R'} + n C_{ox}' \phi_t \right], \quad (5.2.15)$$

which is similar to the conventional expression for the charge in strong inversion, except for the rightmost term. We can rewrite (5.2.15) as an explicit function of the inversion charge densities at source and drain as

$$Q_I = WL \frac{2/3 (Q_{IS}'^2 + Q_{IS}' Q_{ID}' + Q_{ID}'^2) - n C_{ox}' \phi_t (Q_{IS}' + Q_{ID}')}{Q_{IS}' + Q_{ID}' - 2n C_{ox}' \phi_t}. \quad (5.2.16)$$

The expressions above give the total channel charge in all inversion regions. Let us now interpret two very simple cases of transistor operation. Deep in strong inversion, the terms in (5.2.16) containing the thermal charge $n C_{ox}' \phi_t$ are negligible and in saturation $Q_{ID}' \cong 0$. Consequently, $Q_I = (2/3) WL Q_{IS}'$, which is the classical result for the charge in SI saturation [6]. Deep in weak inversion, the quadratic terms are negligible because $|Q_I'| \ll n C_{ox}' \phi_t$. Thus, (5.2.16) reduces to

$$Q_I = WL \frac{(Q_{IS}' + Q_{ID}')}{2}. \quad (5.2.17)$$

Things are very simple in WI, the total charge is proportional to the average charge density in the channel. In fact, for a diffusion-only current, the gradient of the charge in the channel is constant. Although the explicit expression (5.2.16) is handy for limit case analysis, expression (5.2.15) is very useful for symbolic modeling, as will become clear when calculating the capacitive coefficients.

In order to determine the drain and source charges, the coordinate y must be expressed in terms of the inversion charge density Q_I' . The integration of (5.2.6) from the source ($y = 0$) to an arbitrary point y of the channel leads to

$$y = \frac{\mu_n W}{2n C_{ox}' I_D} (Q_F'^2 - Q_I'^2). \quad (5.2.18)$$

The substitution of (5.2.4), (5.2.6), and (5.2.18) into (5.1.14) yields

$$Q_D = -\frac{\mu_n^2 W^3}{2L(nC'_{ox} I_D)^2} \left[\int_{Q'_F}^{Q'_R} (Q'^2_F - Q'^2_{It}) (Q'_{It} + nC'_{ox} \phi_t) Q'_{It} dQ'_{It} \right]. \quad (5.2.19)$$

After calculating the integral, substituting the current I_D in the denominator using expression (5.2.7) and removing the common factor $(Q'_R - Q'_F)^2$ from both numerator and denominator, we obtain

$$Q_D = WL \left[\frac{6Q'^3_R + 12Q'_F Q'^2_R + 8Q'^2_F Q'_R + 4Q'^3_F}{15(Q'_F + Q'_R)^2} + \frac{n}{2} C'_{ox} \phi_t \right]. \quad (5.2.20)$$

Q_S may be determined in a similar way, or through $Q_S = Q_I - Q_D$, or simply by taking advantage of the source and drain symmetry, as follows

$$Q_S = WL \left[\frac{6Q'^3_F + 12Q'_R Q'^2_F + 8Q'^2_R Q'_F + 4Q'^3_R}{15(Q'_F + Q'_R)^2} + \frac{n}{2} C'_{ox} \phi_t \right]. \quad (5.2.21)$$

Table 5.1 Total inversion, source and drain charges as a function of the inversion charge density at source and the saturation coefficient α .

$Q_I = WL \left[\frac{2}{3} \frac{1 + \alpha + \alpha^2}{1 + \alpha} (Q'_{IS} - nC'_{ox} \phi_t) + nC'_{ox} \phi_t \right] \quad (5.2.22)$
$Q_S = WL \left[\frac{6 + 12\alpha + 8\alpha^2 + 4\alpha^3}{15(1 + \alpha)^2} (Q'_{IS} - nC'_{ox} \phi_t) + \frac{n}{2} C'_{ox} \phi_t \right] \quad (5.2.23)$
$Q_D = WL \left[\frac{4 + 8\alpha + 12\alpha^2 + 6\alpha^3}{15(1 + \alpha)^2} (Q'_{IS} - nC'_{ox} \phi_t) + \frac{n}{2} C'_{ox} \phi_t \right] \quad (5.2.24)$

The expressions for the channel-related charges in terms of the source inversion charge density and the saturation coefficient α are summarized in Table 5.1. As seen in Fig. 5.3, in spite of their complicated aspect, the functions of the parameter α are well-behaved, slightly varying functions in the physically meaningful interval $0 < \alpha \leq 1$. These functions give, for a nonuniform channel ($0 < V_{DS}$, or $\alpha < 1$), the average values of the inversion charge density in the channel,

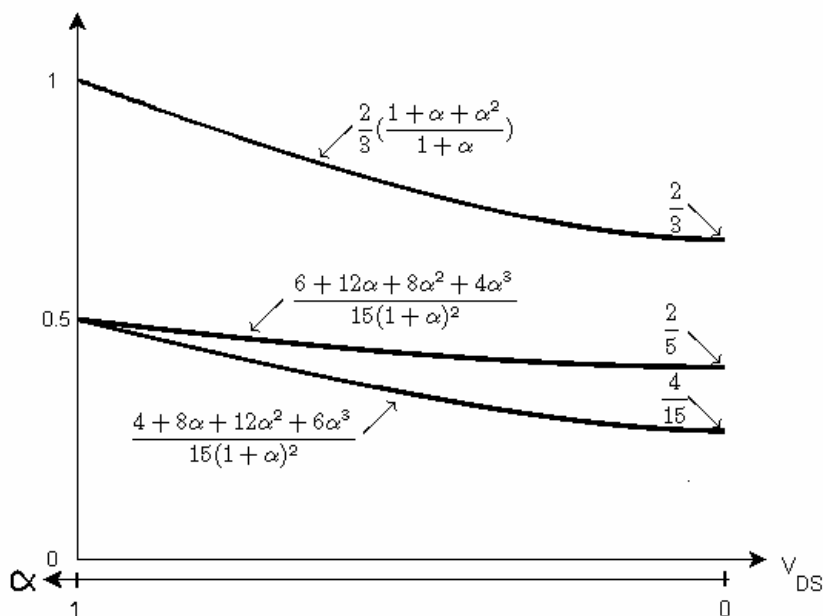


Fig. 5.3 Functions of the inversion coefficient α appearing in the channel charge expressions. (After [6].)

An approximate expression for the depletion charge density, linear in terms of the inversion charge density, is readily obtained using the linear relationship between surface potential and inversion charge in (2.4.41) to express ϕ_s in terms of Q'_I in the potential balance equation (2.3.34), yielding

$$Q'_B \cong -Q'_I - C'_{ox}(V_G - V_{FB} - \phi_{sa} - \frac{Q'_I}{nC'_{ox}}) \quad (5.2.25)$$

or

$$Q'_B \cong -\frac{n-1}{n} Q'_I + Q'_{Ba} \quad (5.2.26)$$

where

$$Q'_{Ba} = -\gamma C'_{ox} \sqrt{\phi_{sa} - \phi_t} \quad (5.2.27)$$

is the depletion charge density deep in WI (if V_C tends toward infinity).

The total depletion charge is directly determined from the definition (5.1.19) using (5.2.26), resulting in

$$Q_B = -\frac{n-1}{n} Q_I + Q'_{Ba} WL. \quad (5.2.28)$$

A more accurate expression for the bulk charge can be deduced (see problem 12.1), but the expression above is useful to derive approximate design-oriented expressions for the bulk capacitances.

The gate charge is determined from the charge neutrality condition

$$Q_G = -Q_B - Q_I - Q_O \quad (5.2.29)$$

where Q_O is the effective oxide charge, assumed to be independent of the terminal voltages. The total stored charges as functions of the gate bias are shown in Fig. 5.4.

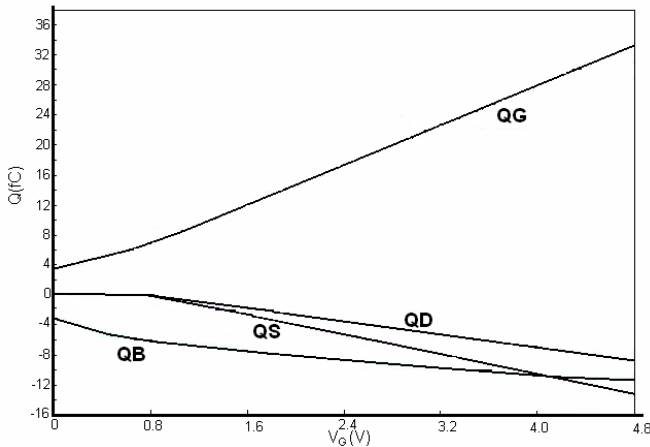


Fig. 5.4 The charges at the gate (Q_G), source (Q_S), drain (Q_D) and bulk (Q_B) terminals versus V_G for an NMOS transistor with $W/L=10\mu\text{m}/0.8\mu\text{m}$. ($V_D=2\text{V}$, $V_S=0\text{V}$).

5.2.2 Transit time

The transit time of the electrons from the source to the drain in an n-channel transistor is readily calculated assuming a constant current along the channel, as is common in compact modeling. By definition, the velocity of the carriers is related to the drain current by

$$I_D = -WQ'_I v_{elec}. \quad (5.2.30)$$

The transit time for a channel element of length Δy is given by

$$\Delta t = \frac{\Delta y}{v_{elec}}. \quad (5.2.31)$$

Thus, the channel transit time τ is

$$\tau = \int_0^L \frac{dy}{v_{elec}} = \int_0^L \frac{dy}{-\frac{I_D}{WQ'_I}} = -\frac{W}{I_D} \int_0^L Q'_I dy = -\frac{Q_I}{I_D}. \quad (5.2.32)$$

Using equation (5.2.15) for the total channel charge and equation (5.2.7) for the drain current, a general expression for the transit time, valid from weak to strong inversion and from the linear to the saturation region is obtained. For the saturated transistor this general expression simplifies to

$$\tau = \frac{2L^2}{\mu\phi_t} \frac{1 + (2/3)q'_{IS}}{(2 + q'_{IS})^2} \quad (5.2.33)$$

where $q'_{IS} = Q'_{IS} / (-nC'_{ox}\phi_t)$ is the normalized inversion charge at the source. In the limit cases of weak (WI) and strong inversion (SI), equation (5.2.33) reduces to the conventional expressions

$$\tau \cong \frac{L^2}{2\mu\phi_t} \text{ (WI)} \quad (5.2.34)$$

and

$$\tau = \frac{4}{3} \frac{nL^2}{\mu} \frac{1}{V_G - V_{T0}} \text{ (SI)}. \quad (5.2.35)$$

(5.2.35) is valid for $V_{SB}=0V$. The transit time is useful as an estimate of the charging time of the transistor channel, which gives the intrinsic delay of the transistor.

5.2.3 Capacitive coefficients

We will first consider the intrinsic part of the transistor. The extrinsic part can be modeled by a network of parasitic (two terminal) capacitances, resistances and diodes.

Since using the quasi-static approximation we express all charges Q_j ($j=G, S, D, B$) in terms of the instantaneous values of the terminal voltages, by applying the chain rule of differentiation we obtain the expression of the charging currents as

$$\frac{dQ_j}{dt} = \frac{\partial Q_j}{\partial V_G} \frac{dV_G}{dt} + \frac{\partial Q_j}{\partial V_S} \frac{dV_S}{dt} + \frac{\partial Q_j}{\partial V_D} \frac{dV_D}{dt} + \frac{\partial Q_j}{\partial V_B} \frac{dV_B}{dt}. \quad (5.2.36)$$

The four-by-four matrix of the MOSFET intrinsic capacitances for quasi-static operation is defined by:

$$C_{jk} = - \left. \frac{\partial Q_j}{\partial V_k} \right|_0 \quad j \neq k \quad (5.2.37)$$

$$C_{jj} = \left. \frac{\partial Q_j}{\partial V_j} \right|_0, \quad (5.2.38)$$

where Q_j can be any of the charges Q_S, Q_D, Q_B , or Q_G and V_j and V_k can be any of the voltages V_S, V_D, V_B , or V_G . The notation “0” indicates that the derivatives are calculated at the bias point. The sign choice in (5.2.37) is standard, and with it most of the capacitive coefficients (but not all of them) are positive. The term C_{jk} can be interpreted as a mutual capacitance. C_{jk} determines the current transferred out of node j because of a voltage change on node k , all the other node voltages remaining constant.

Because the MOSFET is an active device, the capacitances C_{jk} are non-reciprocal, that is, in general, $C_{jk} \neq C_{kj}$ as will be clear after the analytical calculation of the capacitive coefficients of the MOSFET.

Writing the charge derivatives in (5.2.36) as the capacitive coefficients, the charging currents become:

$$\begin{pmatrix} dQ_G / dt \\ dQ_S / dt \\ dQ_D / dt \\ dQ_B / dt \end{pmatrix} = \begin{pmatrix} C_{gg} & -C_{gs} & -C_{gd} & -C_{gb} \\ -C_{sg} & C_{ss} & -C_{sd} & -C_{sb} \\ -C_{dg} & -C_{ds} & C_{dd} & -C_{db} \\ -C_{bg} & -C_{bs} & -C_{bd} & C_{bb} \end{pmatrix} \begin{pmatrix} dV_G / dt \\ dV_S / dt \\ dV_D / dt \\ dV_B / dt \end{pmatrix}. \quad (5.2.39)$$

The 16 coefficients are not linearly independent as will be shown next. Assume

$$V_G(t) = V_S(t) = V_D(t) = V_B(t) = V(t). \quad (5.2.40)$$

Since the voltage differences between transistor nodes remain constant, the charging currents are zero. For the gate charging current we have

$$\frac{dQ_G}{dt} = (C_{gg} - C_{gs} - C_{gd} - C_{gb}) \frac{dV}{dt} = 0. \quad (5.2.41)$$

Thus,

$$C_{gg} = C_{gs} + C_{gd} + C_{gb}. \quad (5.2.42)$$

The same reasoning can be applied to the three remaining rows of the matrix, leading us to conclude that the coefficients of each row of the capacitive matrix must add up to zero.

Let us now assume that

$$\frac{dV_S}{dt} = \frac{dV_D}{dt} = \frac{dV_B}{dt} = 0. \quad (5.2.43)$$

Applying (5.2.39) we obtain

$$\begin{aligned} \frac{dQ_G}{dt} &= C_{gg} \frac{dV_G}{dt}, & \frac{dQ_S}{dt} &= -C_{sg} \frac{dV_G}{dt}, \\ \frac{dQ_D}{dt} &= -C_{dg} \frac{dV_G}{dt}, & \frac{dQ_B}{dt} &= -C_{bg} \frac{dV_G}{dt}. \end{aligned} \quad (5.2.44)$$

The sum of all the charging currents is

$$\frac{dQ_G}{dt} + \frac{dQ_S}{dt} + \frac{dQ_D}{dt} + \frac{dQ_B}{dt} = (C_{gg} - C_{sg} - C_{dg} - C_{bg}) \frac{dV_G}{dt}. \quad (5.2.45)$$

Taking into account charge conservation, $d(Q_S + Q_D + Q_B + Q_G)/dt = 0$, it follows that

$$C_{gg} = C_{sg} + C_{dg} + C_{bg}. \quad (5.2.46)$$

Repeating the same procedure for each node voltage it follows that, as a consequence of charge conservation, each column of the capacitance matrix must add up to zero.

Finally, we have the following linear relationship between capacitive coefficients

$$\begin{aligned} C_{gg} &= C_{gs} + C_{gd} + C_{gb} = C_{sg} + C_{dg} + C_{bg} \\ C_{ss} &= C_{sg} + C_{sd} + C_{sb} = C_{gs} + C_{ds} + C_{bs} \\ C_{dd} &= C_{dg} + C_{ds} + C_{db} = C_{gd} + C_{sd} + C_{bd} \\ C_{bb} &= C_{bg} + C_{bs} + C_{bd} = C_{gb} + C_{sb} + C_{db}. \end{aligned} \quad (5.2.47)$$

Owing to the above relationships, only nine out of the sixteen capacitive coefficients are linearly independent. In effect, each row is obtained from the sum of the other three while the same result applies to each column.

Summarizing, it can be said that due to charge conservation and the fact that only three voltage differences out of four can be chosen independently, the MOSFET capacitive model is characterized by nine independent capacitances. Clearly, in the equivalent circuit of the transistor, we cannot associate the capacitive coefficients to two terminal capacitors only, because we cannot connect more than six capacitors between four nodes. Equivalent electrical circuits for the transistor will be discussed in the next section.

5.2.3.1 Intrinsic small-signal models

Many different equivalent circuits can be derived to represent equation (5.2.39). To derive such a circuit we must first make the choice of a set of nine independent capacitances to represent the complete set of sixteen

capacitive coefficients. The choice of C_{gs} , C_{gd} , C_{gb} , C_{bs} , and C_{bd} is natural because these capacitances have been widely used in ac modeling. There are two reasons for their use: the five above capacitances together describe accurately charge storage in MOSFETs up to moderate frequencies and can be calculated directly from the gate and bulk charges [6].

To preserve the symmetry of the device in the small-signal schematic, an appropriate choice would be to consider C_{ds} , C_{sd} , C_{sg} , and C_{dg} to complete the nine independent capacitance set. For reasons that will become clear in the next section, a non symmetric (regarding source and drain terminals) set of capacitances will be adopted: C_{gs} , C_{gd} , C_{gb} , C_{bg} , C_{bs} , C_{bd} , C_{ds} , C_{sd} , and C_{dg} .

Calculating in (5.2.39) all the capacitive coefficients in terms of the nine capacitances adopted, using (5.2.47), we obtain

$$\begin{aligned}\frac{dQ_G}{dt} &= C_{gs} \frac{dV_{GS}}{dt} + C_{gd} \frac{dV_{GD}}{dt} + C_{gb} \frac{dV_{GB}}{dt} \\ \frac{dQ_B}{dt} &= C_{gb} \frac{dV_{BG}}{dt} + C_{mx} \frac{dV_{BG}}{dt} + C_{bs} \frac{dV_{BS}}{dt} + C_{bd} \frac{dV_{BD}}{dt} \\ \frac{dQ_D}{dt} &= C_{gd} \frac{dV_{DG}}{dt} + C_{bd} \frac{dV_{DB}}{dt} - C_m \frac{dV_{GB}}{dt} + C_{sd} \frac{dV_{DB}}{dt} \\ &\quad - C_{ds} \frac{dV_{SB}}{dt}\end{aligned}\tag{5.2.48}$$

where

$$\begin{aligned}C_m &= C_{dg} - C_{gd} \\ C_{mx} &= C_{bg} - C_{gb}.\end{aligned}\tag{5.2.49}$$

The equivalent circuit representing the equations in (5.2.48) is shown in Fig. 5.5. With the exception of the source-drain branch, the circuit of Fig. 5.5 is identical to that of page 449 of Tsividis' book [6]. In the equivalent circuit of Fig. 5.5, the source-drain current sources are controlled by the gate-to-bulk, source-to-bulk, and drain-to-bulk voltages while in Tsividis' book, the current sources connected between drain and source are controlled by the gate-to-source, drain-to-source, and bulk-to-

source voltages as control voltages. Clearly, the two circuits are strictly equivalent, as the reader can easily verify.

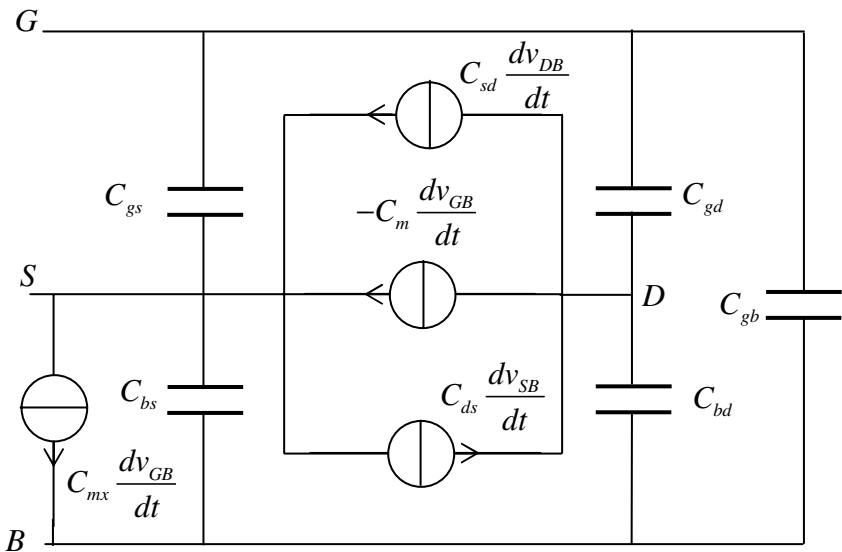


Fig. 5.5 Quasi-static small-signal model for the charging currents of the MOS transistor.

5.2.3.2 Calculation of the capacitive coefficients [9]

From the definitions of capacitive coefficients in (5.2.37) and (5.2.38), and from the expressions of the MOSFET charges presented in Section 5.2.1, we can derive formulas for the small-signal capacitances.

Substituting the expression for the bulk charge, (5.2.28), into the charge balance equation, (5.2.29), we obtain:

$$Q_G = -\frac{Q_I}{n} - Q'_{Ba} WL - Q_O. \quad (5.2.50)$$

Capacitances C_{gs} and C_{gd} can be obtained as follows. Using definition (5.2.37) and deriving the gate charge (5.2.50) with respect to $V_{S(D)}$ yields

$$C_{gs} = -\frac{\partial Q_G}{\partial V_S} = \frac{1}{n} \frac{\partial Q_I}{\partial V_S} \quad (5.2.51)$$

and

$$C_{gd} = -\frac{\partial Q_G}{\partial V_D} = -\frac{1}{n} \frac{\partial Q_I}{\partial V_D} \quad (5.2.52)$$

because Q'_{Ba} and Q_O do not depend on either V_S or V_D . Substituting expression (5.2.15) for the total inversion charge Q_I into (5.2.51) and (5.2.52) gives

$$C_{gs} = \frac{1}{n} \frac{\partial Q_I}{\partial Q'_F} \frac{\partial Q'_F}{\partial V_S} = \frac{2WL}{3n} \left[\frac{Q'^2_F + 2Q'_F Q'_R}{(Q'_F + Q'_R)^2} \right] \frac{\partial Q'_{IS}}{\partial V_S} \quad (5.2.53)$$

and

$$C_{gd} = \frac{1}{n} \frac{\partial Q_I}{\partial Q'_R} \frac{\partial Q'_R}{\partial V_D} = \frac{2WL}{3n} \left[\frac{Q'^2_R + 2Q'_R Q'_F}{(Q'_F + Q'_R)^2} \right] \frac{\partial Q'_{ID}}{\partial V_D}. \quad (5.2.54)$$

From (2.4.57), the equation for the derivative of the source (drain) inversion charge with respect to the source (drain) voltage is

$$\frac{\partial Q'_{IS(D)}}{\partial V_{S(D)}} = nC'_{ox} \frac{Q'_{IS(D)}}{Q'_{IS(D)} - nC'_{ox} \phi_t}. \quad (5.2.55)$$

The substitution of (5.2.55) into (5.2.53) and (5.2.54) results in

$$C_{gs(d)} = \frac{2}{3} C_{ox} \left[1 - \frac{Q'^2_{R(F)}}{(Q'_F + Q'_R)^2} \right] \left[1 + \frac{nC'_{ox} \phi_t}{Q'_{F(R)}} \right] \quad (5.2.56)$$

where $C_{ox} = WLC'_{ox}$ is the total intrinsic oxide capacitance.

We can calculate the source-to-drain capacitance C_{sd} by directly calculating the derivative of the total source charge, given by (5.2.21), with respect to V_D and by using (5.2.55) to calculate the derivative of the inversion charge. Thus

$$C_{sd} = -\frac{4}{15} nC_{ox} \frac{Q'^3_R + 3Q'^2_R Q'_F + Q'_R Q'^2_F}{(Q'_F + Q'_R)^3} \left(1 + \frac{nC'_{ox} \phi_t}{Q'_R} \right). \quad (5.2.57)$$

Taking the derivative of the total drain charge with respect to V_S we obtain the drain-to-source capacitance C_{ds}

$$C_{ds} = -\frac{4}{15} n C_{ox} \frac{Q_F'^3 + 3Q_F'^2 Q_R' + Q_F' Q_R'^2}{(Q_F' + Q_R')^3} \left(1 + \frac{n C_{ox}' \phi_t}{Q_F'} \right). \quad (5.2.58)$$

Clearly, we can obtain (5.2.58) from (5.2.57) invoking the symmetry between source and drain.

Using expression (5.2.28) for the bulk charge, the bulk-to-source C_{bs} and the bulk-to-drain C_{bd} capacitances are given by

$$C_{bs} = -\frac{\partial Q_B}{\partial V_S} \bigg|_{V_D, V_G, V_B} = \frac{n-1}{n} \frac{\partial Q_I}{\partial V_S} \bigg|_{V_D, V_G, V_B} \quad (5.2.59)$$

$$C_{bd} = -\frac{\partial Q_B}{\partial V_D} \bigg|_{V_S, V_G, V_B} = \frac{n-1}{n} \frac{\partial Q_I}{\partial V_D} \bigg|_{V_S, V_G, V_B}. \quad (5.2.60)$$

The comparison of (5.2.51) and (5.2.52) with (5.2.59) and (5.2.60) allows us to write

$$C_{bs} = (n-1)C_{gs} \quad (5.2.61)$$

$$C_{bd} = (n-1)C_{gd}. \quad (5.2.62)$$

To obtain the gate-to-bulk C_{gb} and the bulk-to-gate C_{bg} capacitances, the derivatives of charges and voltages summarized in Table 5.2 are helpful.

Using (5.2.69) and neglecting the variation of n with V_G , it follows from expressions (5.2.50) and (5.2.28) for charges Q_G and Q_B , respectively, that:

$$C_{gb} = \frac{1}{n} \frac{\partial Q_I}{\partial V_B} \bigg|_{V_S, V_D, V_G} + \frac{n-1}{n} C_{ox} \quad (5.2.63)$$

$$C_{bg} = \frac{n-1}{n} \frac{\partial Q_I}{\partial V_G} \bigg|_{V_S, V_D, V_B} + \frac{n-1}{n} C_{ox}. \quad (5.2.64)$$

Applying the chain rule and (5.2.67) to calculate the partial derivatives of Q_I yields

$$\frac{\partial Q_I}{\partial V_G} = -\frac{1}{n} \frac{\partial Q_I}{\partial Q'_F} \frac{\partial Q'_{IS}}{\partial V_S} - \frac{1}{n} \frac{\partial Q_I}{\partial Q'_R} \frac{\partial Q'_{ID}}{\partial V_S}. \quad (5.2.65)$$

Table 5.2 Relationships between the partial derivatives of charge densities.

$\frac{\partial Q'_I}{\partial V_C} = -\frac{\partial Q'_I}{\partial V_P}$	(5.2.66)
$\frac{\partial Q'_I}{\partial V_G} = -\frac{1}{n} \frac{\partial Q'_I}{\partial V_C}$	(5.2.67)
$\frac{\partial Q'_I}{\partial V_B} = -\frac{n-1}{n} \frac{\partial Q'_I}{\partial V_C}$	(5.2.68)
$\left. \frac{\partial Q'_{Ba}}{\partial V_G} \right _{V_B} = -\left. \frac{\partial Q'_{Ba}}{\partial V_B} \right _{V_G} = -\frac{n-1}{n} C'_{ox}$	(5.2.69)
$\frac{\partial V_P}{\partial V_G} = -\frac{\partial V_P}{\partial V_B} = \frac{1}{n}$	(5.2.70)

In a similar way and using (5.2.68) instead of (5.2.67) we obtain

$$\frac{\partial Q_I}{\partial V_B} = -\frac{n-1}{n} \frac{\partial Q_I}{\partial Q'_F} \frac{\partial Q'_{IS}}{\partial V_S} - \frac{n-1}{n} \frac{\partial Q_I}{\partial Q'_R} \frac{\partial Q'_{ID}}{\partial V_D}. \quad (5.2.71)$$

Comparing the results above with expressions (5.2.53) and (5.2.54) for C_{gs} and C_{gd} , it follows that

$$\begin{aligned} \frac{\partial Q_I}{\partial V_G} &= -(C_{gs} + C_{gd}) \\ \frac{\partial Q_I}{\partial V_B} &= -(n-1)(C_{gs} + C_{gd}). \end{aligned} \quad (5.2.72)$$

Finally, from (5.2.68), (5.2.69), and (5.2.72) we obtain:

$$C_{gb} = C_{bg} = \frac{n-1}{n} (C_{ox} - C_{gs} - C_{gd}). \quad (5.2.73)$$

As a consequence of the approximations made, we obtain $C_{gb}=C_{bg}$. However, accurate calculations [7] carried out using the charge sheet model give

$$C_{gb} < C_{bg} \quad (5.2.74)$$

and, consequently, $C_{mx}>0$. Due to its very small value C_{mx} is considered unimportant in most practical cases [6].

Calculating the partial derivative of Q_S with respect to V_G using (5.2.67) and comparing the result thus obtained with the expressions for C_{ss} and C_{sd} gives

$$C_{sg} = (C_{ss} - C_{sd}) / n. \quad (5.2.75)$$

A similar expression is obtained for C_{gs} :

$$C_{gs} = (C_{ss} - C_{ds}) / n. \quad (5.2.76)$$

Finally, in the case $C_{gb}=C_{bg}$ it follows from (5.2.47) that

$$C_m = C_{dg} - C_{gd} = C_{gs} - C_{sg}. \quad (5.2.77)$$

Combining (5.2.77) with (5.2.75) and (5.2.76) yields

$$C_m = (C_{sd} - C_{ds}) / n. \quad (5.2.78)$$

The small-signal schematic of Fig. 5.6, comprising five capacitances, three transcapacitances, and three transconductances preserves the inherent symmetry of the MOSFET. The use of the normalized inversion charges $q'_{IS(D)} = Q'_{IS(D)} / (-nC'_{ox}\phi_t)$ and of the saturation coefficient $\alpha = Q'_R / Q'_F$ gives us the meaningful expressions for the capacitive coefficients in Table 5.3. The symmetry between the source and drain terminals is apparent in the formulas of Table 5.3. Exchanging source and drain terminals corresponds to exchanging α for $1/\alpha$ and $q'_{IS(D)}$ for $q'_{ID(S)}$. For $V_{DS}=0$, $q'_{IS} = q'_{ID}$ and $\alpha=1$; consequently $C_{sd}=C_{ds}$. It can be

easily verified that this result holds for all the capacitive coefficients at $V_{DS}=0$, thus $C_{jk}=C_{kj}$ for $V_{DS}=0$. However, the capacitive coefficients are nonreciprocal in the general case.

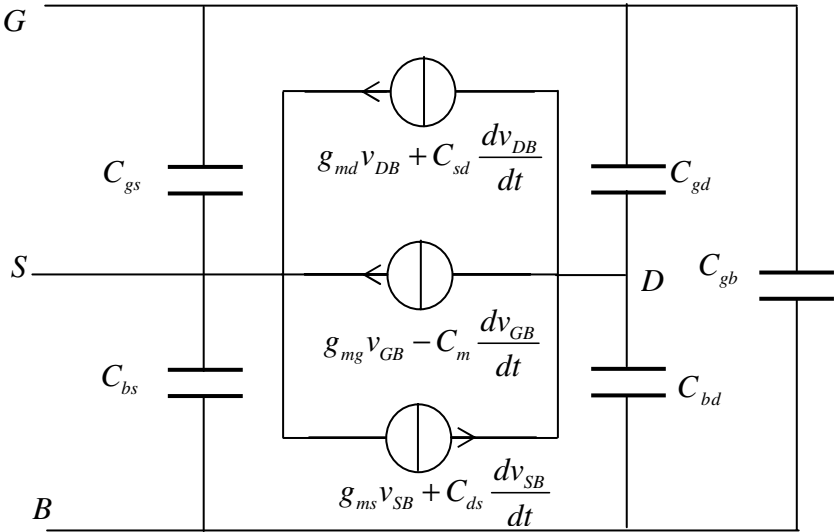


Fig. 5.6 Simplified small signal MOSFET model .

As an example let us consider C_{sd} and C_{ds} in saturation ($q'_{ID}=0$). From the equations in Table 5.3, it follows that $C_{sd}=0$, but $C_{ds}\neq 0$. In saturation, the (long-channel) transistor charges are independent of V_D , consequently $C_{sd}=-\partial Q_S/\partial V_D=0$, but since Q_D depends on V_S then $C_{ds}=-\partial Q_D/\partial V_S\neq 0$.

The asymptotic cases of strong and weak inversion are readily obtained from the formulas of Table 5.3. In SI, $q'_{IS(D)} \gg 1$; consequently $q'_{IS(D)} \gg q'_{IS(D)} / (1 + q'_{IS(D)}) \cong 1$, and we re-encounter the conventional expressions for the capacitive coefficients in SI. In WI $\alpha \cong 1$, $1 + q'_{IS(D)} \cong 1$, and we obtain, for example, $C_{gs(d)} = C_{ox} q'_{IS(D)} / 2$.

Table 5.3 A complete set of 9 capacitive coefficients for the MOSFET.

$C_{gs} = \frac{2}{3} C_{ox} \frac{1+2\alpha}{(1+\alpha)^2} \frac{q'_{IS}}{1+q'_{IS}}$
$C_{gd} = \frac{2}{3} C_{ox} \frac{\alpha^2+2\alpha}{(1+\alpha)^2} \frac{q'_{ID}}{1+q'_{ID}}$
$C_{bs(d)} = (n-1)C_{gs(d)}$
$C_{bg} = C_{gb} = \frac{n-1}{n} (C_{ox} - C_{gs} - C_{gd})$
$C_{sd} = -\frac{4}{15} n C_{ox} \frac{\alpha+3\alpha^2+\alpha^3}{(1+\alpha)^3} \frac{q'_{ID}}{1+q'_{ID}}$
$C_{ds} = -\frac{4}{15} n C_{ox} \frac{1+3\alpha+\alpha^2}{(1+\alpha)^3} \frac{q'_{IS}}{1+q'_{IS}}$
$C_{dg} - C_{gd} = C_m = (C_{sd} - C_{ds}) / n$

In Fig. 5.7 we compare five capacitances calculated from the expressions presented in Table 5.3 and from the charge-sheet ϕ_S -formulated model [7]. The accuracy of the gate-related capacitances is excellent. The errors in the bulk capacitances exhibit maximum values deep in strong inversion owing to the linear approximation of the depletion charge density, (5.2.26). The maximum errors verified in the bulk capacitances are less than $0.1C_{ox}$, and do not affect significantly the behavior of most circuits since these capacitances are generally either in parallel with larger capacitances or even short-circuited.

The shapes of the curves of the capacitances as a function of the gate-to-bulk voltage in Fig. 5.7 can be understood as follows. For low-values of V_G , the only relevant capacitance is C_{gb} , because the channel charge is very small and consequently capacitances C_{gs} and C_{gd} are negligible. At the transition between weak and strong inversion C_{gs} increases sharply

because the channel charge increases rapidly with the gate voltage and becomes predominant over the depletion charge. Capacitance C_{gd} remains negligible since the transistor operates in saturation and the drain voltage has no significant effect on the channel charge. Finally, for high values of the gate voltage the transistor enters the linear region and the two capacitances C_{gs} and C_{gd} approach asymptotically the same value of one half of the total gate capacitance. In strong inversion C_{gb} is negligible because the channel charge screens the bulk from the variation in the gate voltage.

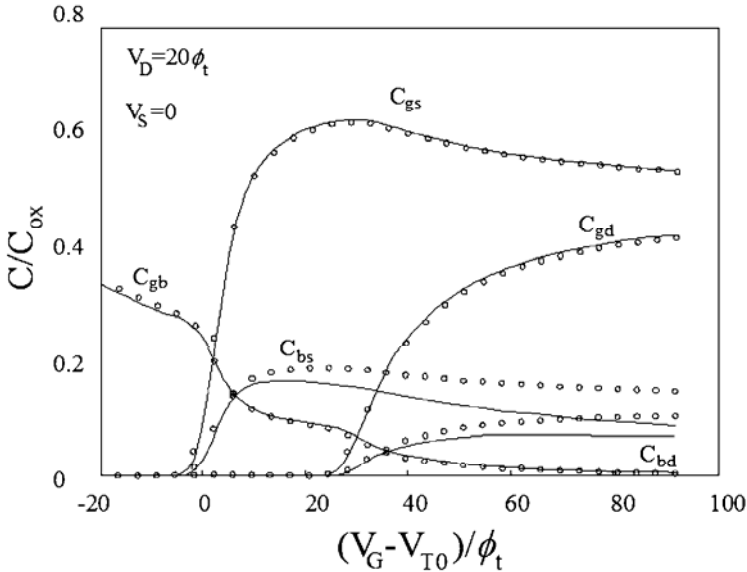


Fig. 5.7 Intrinsic capacitances simulated (—) from the quasi-static model described in Table 5.3 and (o) from the ϕ_5 -formulated model of [7] (NMOS transistor, $t_{ox} = 250 \text{ \AA}$, $N_A = 2 \times 10^{16} \text{ cm}^{-3}$, and $V_{T0} = 0.7 \text{ V}$). (After [8].)

The source-related capacitances (C_{ss} , C_{sg} , C_{sb} and C_{sd}) are shown in Fig. 5.8. As previously commented on, $C_{sd} = 0$ in saturation. Out of saturation the negative value of C_{sd} is easily understood. An increase in the drain voltage ($\Delta V_D > 0$) decreases the magnitude of the (negative) inversion and drain charges. Thus, $\Delta Q_S > 0$ and, consequently, $C_{sd} < 0$.

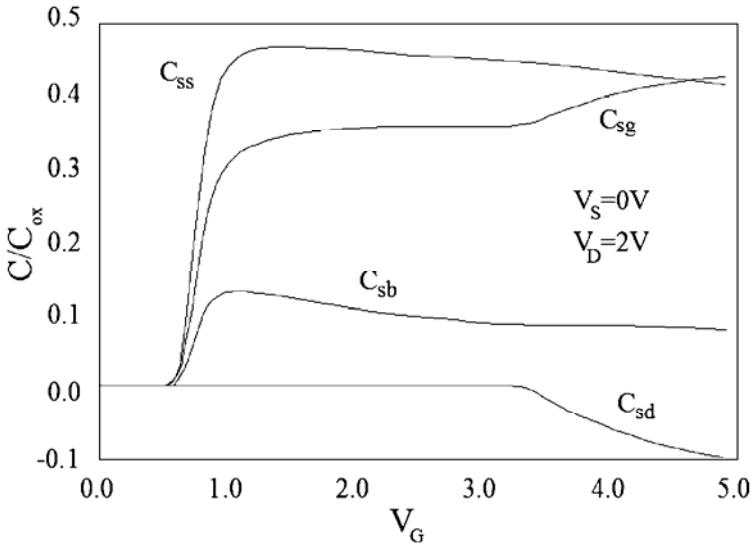


Fig. 5.8 Normalized source-related capacitances. (After [9].)

5.2.3.3 Charge conservation and transcapacitances

The classical 5-capacitance MOSFET model is shown in Fig. 5.9. Recall, however, that a MOSFET model containing only these five capacitances does not conserve charge, because these five capacitances are unable to represent the stored charges consistently. Since the charges stored in the MOSFET are nonlinear functions of several variables (the node voltages), the 5-capacitor model represents an incomplete subset of the partial derivatives for the calculation of charge variation.

A simple test to verify charge conservation is to simulate the sample-hold circuit shown in Fig. 5.10.

The simulations were performed using the ACM, SPICE3, and BSIM3v3 models. The ACM and BSIM3v3 models use charge-based capacitance models while SPICE3 uses the 5-capacitor model. Fig. 5.11 shows the simulation results. Both ACM and BSIM3v3 give results consistent with the physical behavior while SPICE3 does not. Therefore, for the electrical simulation of charge sensitive circuits such as switched-

capacitor or switched-current filters, the complete quasi-static model must be taken into account [6], [10].

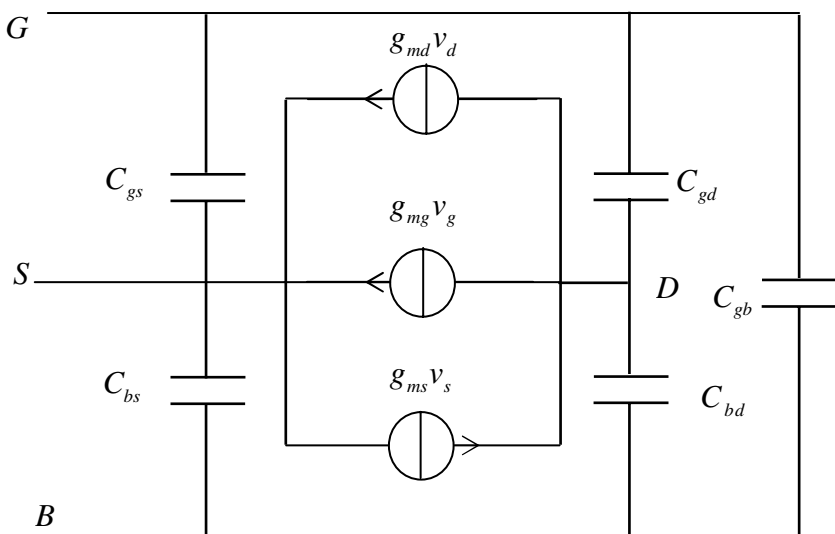


Fig. 5.9 Simple 5-capacitance small-signal equivalent circuit for the MOSFET.

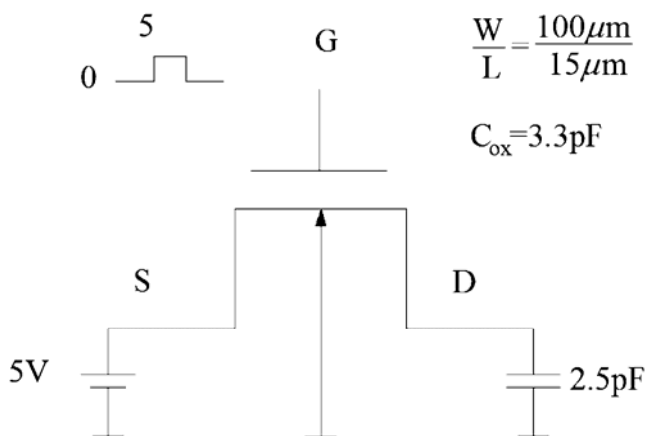


Fig. 5.10 Sample-hold circuit employed to verify charge-conservation.

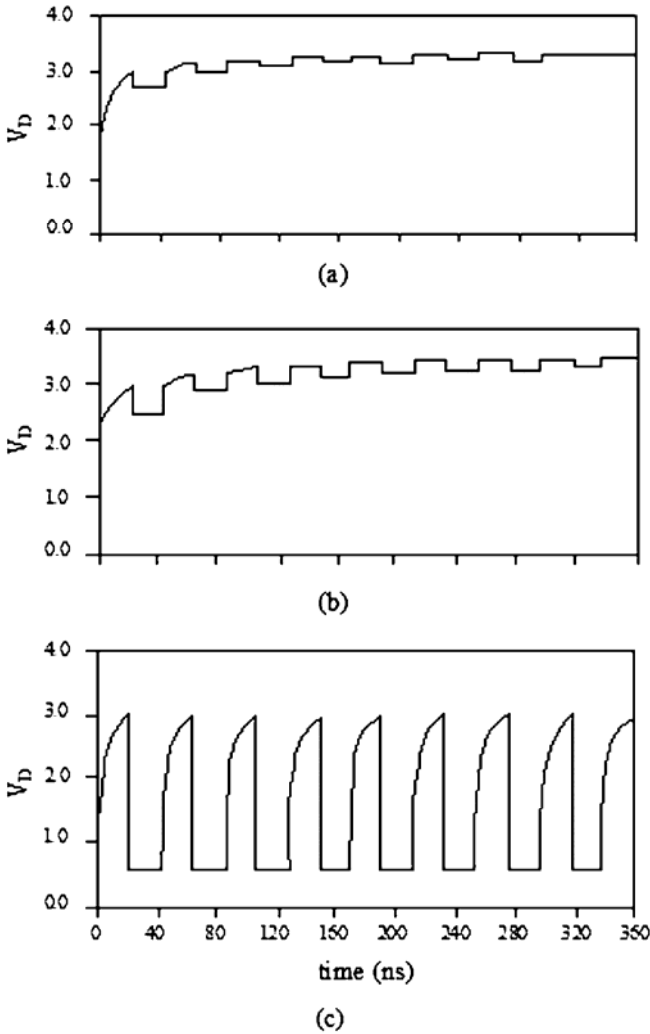


Fig. 5.11 Simulation of the sample-and-hold circuit in Fig. 5.10: (a) ACM model, (b) BSIM 3v3, (c) SPICE3. (After [11].)

5.2.3.4 Intrinsic transition frequency

An important figure of merit for a MOSFET is the unity-gain or intrinsic cut-off frequency, defined as the frequency value at which the short-circuit current gain in the common-source configuration of Fig. 5.12

drops to 1 [6]. The equivalent circuit reduces to that of Fig. 5.12, where only three capacitances are represented. In effect, since the only non constant voltage is V_G , only capacitances linked to that node or current sources controlled by V_G may be considered. Transcapacitance C_m was neglected and $C_{gd}=0$ because the transition frequency is defined for the transistor in saturation. Small-signal analysis of the equivalent circuit in Fig. 5.12 gives the intrinsic cut-off frequency of a MOSFET in saturation as

$$f_T = \frac{g_{mg}}{2\pi(C_{gs} + C_{gb})} = \frac{g_{ms}}{2\pi n(C_{gs} + C_{gb})}. \quad (5.2.79)$$

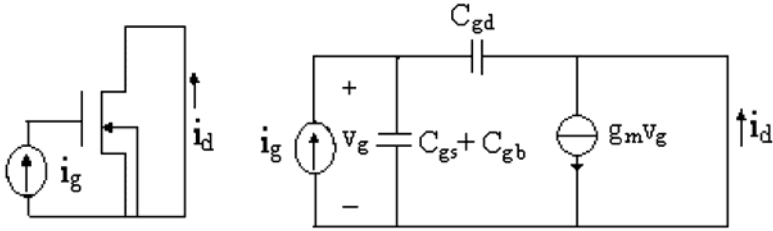


Fig. 5.12 Conceptual circuit to define the intrinsic transition frequency and small-signal equivalent circuit.

From the expressions of the source transconductance g_{ms} given by equation (3.7.26), and the values of C_{gs} and C_{gb} in Table 5.3, f_T can be readily written in terms of the inversion coefficient i_f as:

$$f_T = \frac{\mu n \phi_t}{2\pi L^2} \frac{i_f (\sqrt{1+i_f} + 1)}{(n-1)(\sqrt{1+i_f} + 1)^2 + \frac{2}{3}(i_f + \sqrt{1+i_f} - 1)}. \quad (5.2.80)$$

The first term in the right-hand side of (5.2.80) shows the dependence of f_T on the channel length, slope factor and mobility. The second term represents the dependence of the cut-off frequency on the inversion level.

Sometimes, due to the lack of appropriate models, designers employ transistors whose f_T is much higher than that required for a specific application, thus leading to an unnecessary increase in power consumption.

Assuming the slope factor n in the denominator of (5.2.80) to be equal to $4/3$, a typical value, f_T can be roughly approximated by

$$f_T \cong \frac{\mu\phi_t}{2\pi L^2} 2\left(\sqrt{1+i_f} - 1\right) \quad (5.2.81)$$

for any inversion level.

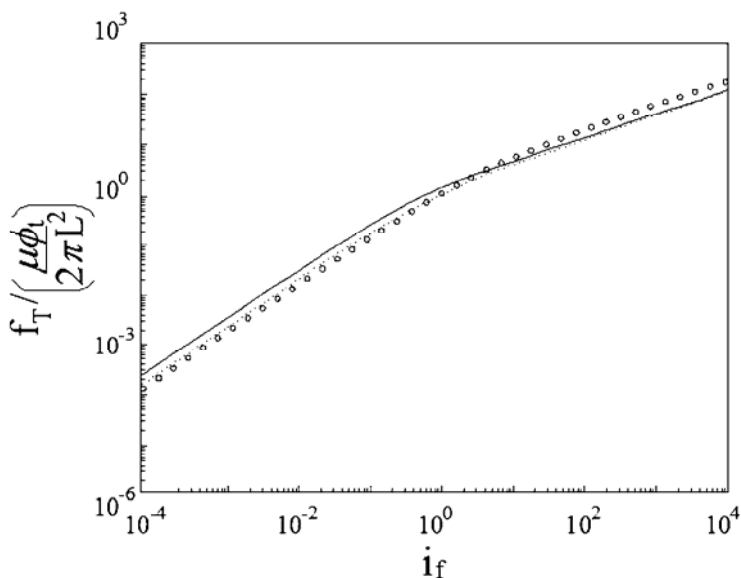


Fig. 5.13 Normalized intrinsic cutoff frequency calculated for (----) $n=4/3$, (—) $n=5/3$ and (o) expression (5.2.81).

Fig. 5.13 shows the intrinsic cut-off frequency calculated from (5.2.80) for $n = 4/3$ and $n = 5/3$, and from (5.2.81) for several decades of the inversion coefficient. Even though (5.2.81) is not very accurate, it gives designers a first order estimate of the unity-gain frequency.

5.3 Charges and capacitances of the extrinsic transistor [6]

The charge storage in the extrinsic parts (Fig. 5.14) of the MOS transistor can be modeled with up to six capacitances, one between each pair of terminals (see Fig. 5.15). The unavoidable overlap between the gate and the source and drain diffusions originates the overlap capacitances. In parallel with the overlap capacitance, the outer fringing and the top capacitances must be included as shown in Fig. 5.16. All three capacitances are approximately proportional to W . Fringing capacitances are particularly important for small dimension devices.

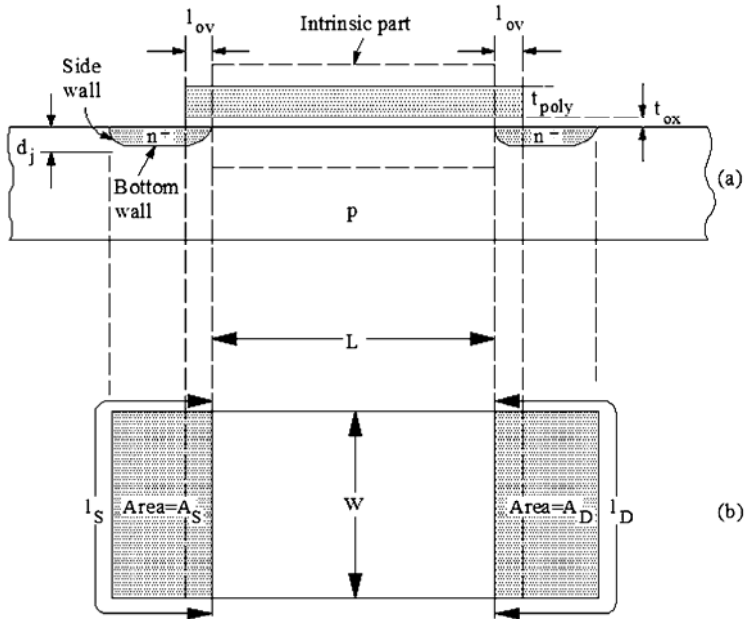


Fig. 5.14 Idealized MOS transistor showing the intrinsic and extrinsic parts. (a) cross-section, (b) top view. (After [6].)

The substrate-source and the substrate-drain junctions must also be modeled by the (nonlinear) diode capacitances. The very small drain-to-source proximity capacitance can usually be neglected and the extrinsic gate-to-bulk capacitance can be incorporated into the gate wiring

capacitance. A more complete model for the extrinsic part should include parasitic resistances as well [6].

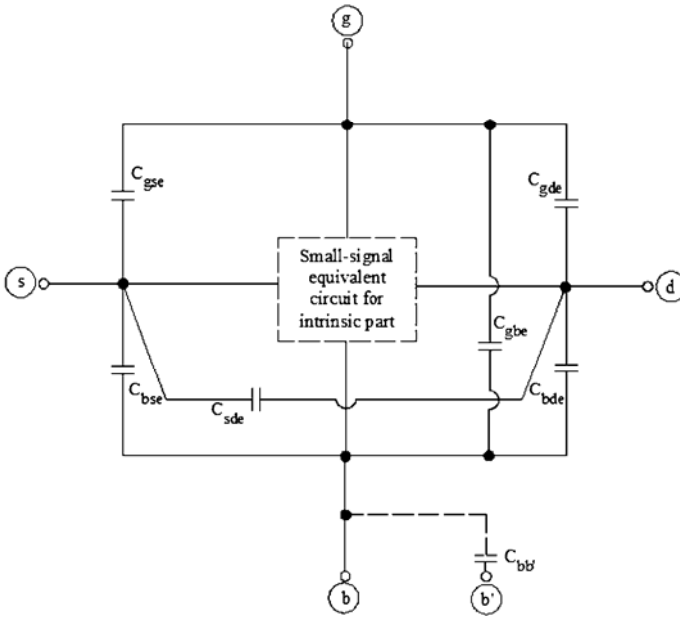


Fig. 5.15 Extrinsic transistor capacitances (After [6]).

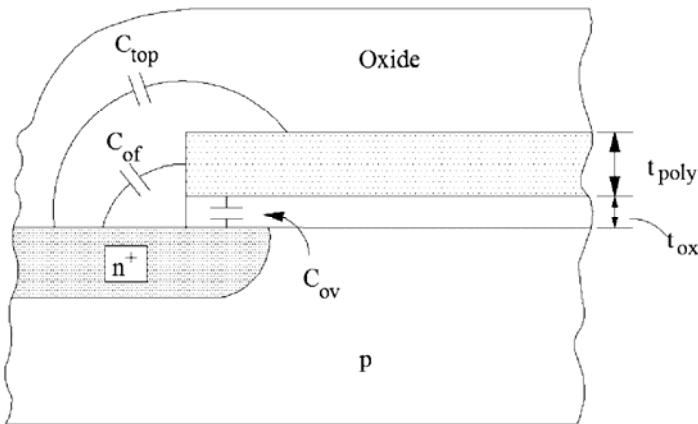


Fig. 5.16 Extrinsic gate-to-diffusion capacitances (After [6]).

5.4 Small-dimension effects on charges and capacitances

5.4.1 Effect of velocity saturation on stored charges

We follow the same procedure as that for the long channel transistor. We change the variable, from length to inversion charge density, to calculate the integrals giving the total charges. To obtain the new relation between the differential of channel length and the differential of inversion charge, we substitute the approximate linear relationship between Q'_I and ϕ_s given by Eq. (3.4.22) into both expression (5.2.1) for the drain current and the field-dependent mobility relation (4.2.3), repeated here for convenience

$$\mu = \frac{\mu_s}{1 + \frac{\mu_s}{v_{sat}} \frac{d\phi_s}{dy}}, \quad (5.4.1)$$

thus obtaining

$$dy = -\frac{\mu_s W}{nC'_{ox} I_D} \left(Q'_I - nC'_{ox} \phi_t + \frac{I_D}{Wv_{sat}} \right) dQ'_I. \quad (5.4.2)$$

Expression (5.4.2) has very important properties. Firstly, because n depends only on V_G , and I_D is constant along the channel, we can define a virtual charge density that differs from the real charge by a constant term, *i.e.*,

$$Q'_V = Q'_I - nC'_{ox} \phi_t + \frac{I_D}{Wv_{sat}}. \quad (5.4.3)$$

The following equality holds along the transistor channel

$$dQ'_V = dQ'_I. \quad (5.4.4)$$

The last two terms in (5.4.3), constant along the channel, have clear physical meanings: $-nC'_{ox} \phi_t$ is the pinch-off charge and $-I_D/Wv_{sat}$ is the saturation charge, *i.e.*, the minimum amount of carrier charge density required to sustain a channel current equal to I_D . Thus, we can rewrite (5.4.3) as

$$Q'_V = Q'_I + Q'_{IP} - Q'_{Isat}. \quad (5.4.5)$$

The virtual charge is the real inversion charge plus the pinch-off charge (diffusion increases the current) minus the saturation charge (the velocity saturation phenomenon reduces the current).

Using (5.4.3) and (5.4.4) we can rewrite (5.4.2) as

$$dy = -\frac{\mu_s W}{nC'_{ox} I_D} Q'_V dQ'_V. \quad (5.4.6)$$

Using the basic linear relationship between Q'_I and ϕ_s given by (3.4.22), Eq. (5.4.6) becomes equivalent to

$$I_D = -\mu_s W Q'_V \frac{d\phi_s}{dy}. \quad (5.4.7)$$

This corresponds to the drift-only current of the virtual charge Q'_V . We can adopt the following definition for the virtual charge: *The drift of the virtual charge produces the same current as the actual movement of the real charge which includes drift, diffusion and velocity saturation.* The use of the virtual charge allows us to extend the formalism of strong inversion to the general case of drift-diffusion plus velocity saturation. The drain current obtained integrating (5.4.6) between source and drain results in

$$I_D = \frac{\mu_s W}{C'_{ox} L} \frac{Q'^2_{VS} - Q'^2_{VD}}{2n}. \quad (5.4.8)$$

Clearly, (5.4.8) is not useful as a stand-alone expression for the calculation of the current because the virtual charge Q'_V depends on the current itself. Nevertheless, the expression of the current in terms of the virtual charges at the source and drain ends is instrumental for deriving simple expressions of the stored charges and capacitive coefficients that keep the drift-only transport formalism.

To calculate the total inversion charge, the channel is split into the saturated and non-saturated regions. In the saturated region the inversion charge density is assumed to be constant; therefore, the inversion charge is given by

$$Q_I = W \int_0^{L-\Delta L} Q'_I dy + W \Delta L Q'_{IDsat}. \quad (5.4.9)$$

To determine the inversion charge in (5.4.9) we calculate Q'_I in terms of Q'_V by means of (5.4.3) and y in terms of Q'_V using (5.4.6). Thus, the total inversion charge is

$$Q_I = W(L - \Delta L) \left[\frac{2}{3} \frac{Q'^2_{VS} + Q'_{VS} Q'_{VD} + Q'^2_{VD}}{Q'_{VS} + Q'_{VD}} + nC'_{ox} \phi_t \right] - \frac{LI_D}{v_{sat}}. \quad (5.4.10)$$

The similarity of the result to the long-channel case is self-evident, clearly showing the importance of the use of the virtual charge as the key variable. In the limit case of velocity saturation along the whole channel, $\Delta L = L$ and the total inversion charge reduces to $-LI_D / v_{sat} = WLQ'_{IDsat}$ as expected.

Neglecting short- and narrow-channel effects on the threshold voltage, the depletion and gate charges are given by the same expressions as in the case of the long-channel transistor

$$Q_B = -\frac{n-1}{n} Q_I - WL \frac{\gamma^2 C'_{ox}}{2(n-1)} \quad (5.4.11)$$

$$Q_G = -Q_B - Q_I. \quad (5.4.12)$$

The drain charge is given by

$$Q_D = W \int_0^{L-\Delta L} \frac{y}{L} Q'_I dy + W \int_{L-\Delta L}^L \frac{y}{L} Q'_{IDsat} dy \quad (5.4.13)$$

where the integrals are calculated along the saturated and nonsaturated regions of the channel. The coordinate y is obtained by integrating (5.4.6) from the source ($y = 0$, $Q'_V = Q'_{VS}$) to an arbitrary point of the channel (y , Q'_V)

$$y = \frac{\mu_s W}{2nC'_{ox} I_D} (Q'^2_{VS} - Q'^2_V). \quad (5.4.14)$$

Substituting (5.4.6) and (5.4.14) into (5.4.13) and expressing Q'_I in terms of Q'_V using (5.4.3), after integration and some algebra (the same as that for the long-channel case) we obtain

$$Q_D = \frac{W(L-\Delta L)^2}{L} \left(\frac{2}{15} \frac{3Q_{VD}'^3 + 6Q_{VD}'^2 Q_{VS}' + 4Q_{VD}' Q_{VS}'^2 + 2Q_{VS}'^3}{(Q_{VS}' + Q_{VD}')^2} + \frac{nC_{ox}' \phi_t}{2} \right) - \frac{LI_D}{2v_{sat}}. \quad (5.4.15)$$

As can be seen in (5.4.15), in the limit case of velocity saturation along the whole channel ($\Delta L=L$), half of the total inversion charge $-LI_D/2v_{sat}$ is associated with the drain.

Finally, the source charge is

$$Q_S = Q_I - Q_D. \quad (5.4.16)$$

The effect of velocity saturation is an increase in the absolute values of the total inversion charges in saturation as a consequence of the non zero charge density at the drain end. This effect is illustrated in Fig. 5.17 where the source and drain charges considering velocity saturation are plotted together with the charges obtained from the long-channel model.

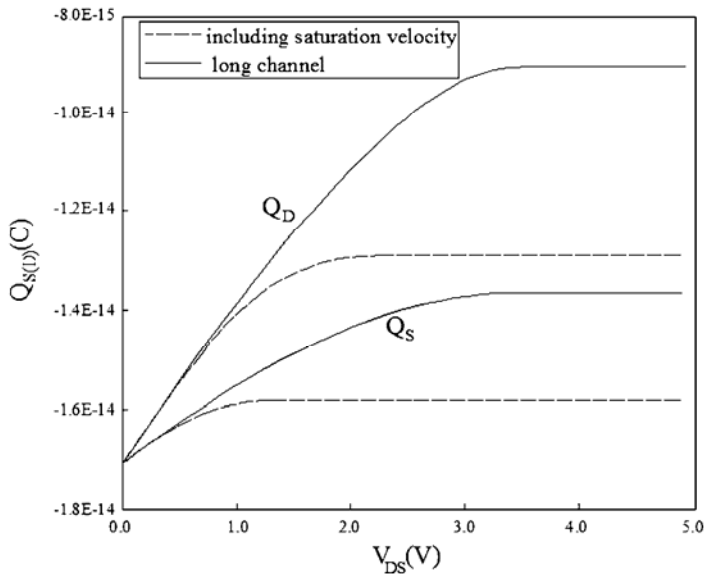


Fig. 5.17 Source and drain charges versus drain voltage for long-channel device and for short-channel device in strong inversion. (After [11].)

Another difference between the long-channel and short-channel characteristics is the reduction in the saturation voltage due to the finite saturation velocity of the carriers.

In Fig. 5.18 the effects of both CLM and DIBL are shown. DIBL causes an increase in the absolute value of the charges, in the saturation region, because of the increase (reduction) of the pinch-off (threshold) voltage with increasing drain voltages. DIBL combined with velocity saturation can produce a maximum in the charge characteristics, as shown in Fig. 5.18 [11].

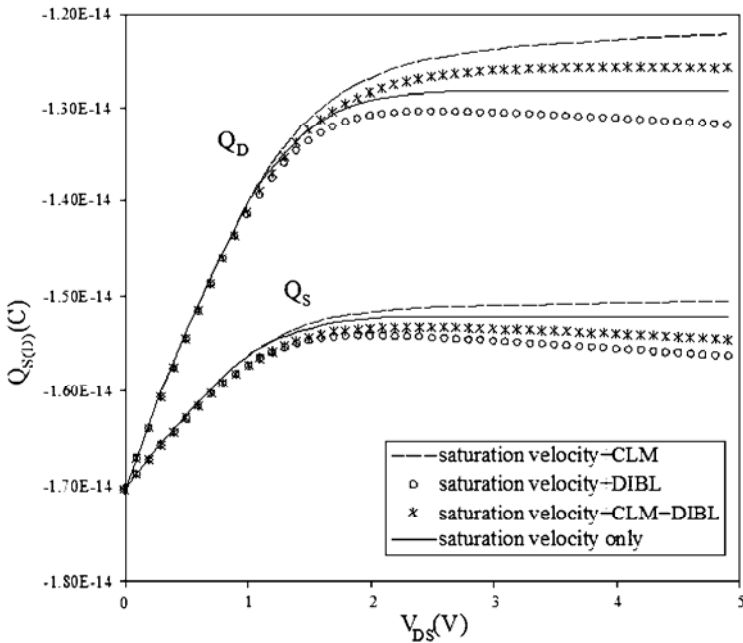


Fig. 5.18 Drain and source charges showing the effects of CLM and DIBL separately and combined with one another. (After [11].)

5.4.2 Effect of velocity saturation on capacitances

Applying the definition (5.2.37) of the capacitive coefficients and using the gate charge expression given in (5.4.12) together with (5.4.11), the

gate to source capacitance C_{gs} is given by the same expression as that for the long-channel transistor, repeated here for convenience

$$C_{gs} = -\frac{\partial Q_G}{\partial V_S} = -\frac{1}{n} \frac{\partial Q_I}{\partial V_S}. \quad (5.4.17)$$

From (5.4.10) we have

$$\frac{\partial Q_I}{\partial V_S} = \frac{\partial Q_I}{\partial Q'_{VS}} \frac{\partial Q'_{VS}}{\partial V_S} + \frac{\partial Q_I}{\partial Q'_{VD}} \frac{\partial Q'_{VD}}{\partial V_S} + \frac{L_e g_{ms}}{v_{sat}} \quad (5.4.18)$$

where $L_e = L - \Delta L$ is the electrical channel length. In (5.4.18) we have assumed that $L \gg \Delta L$. This approximation allows a more compact expression for the capacitance. Also, the partial derivatives of the virtual charges each have an additional term related to the saturation velocity as shown in expressions (5.4.19) and (5.4.20)

$$\frac{\partial Q'_{VS}}{\partial V_S} = \frac{\partial Q'_I}{\partial V_S} - \frac{g_{ms}}{W v_{sat}} \quad (5.4.19)$$

$$\frac{\partial Q'_{VD}}{\partial V_S} = -\frac{g_{ms}}{W v_{sat}}. \quad (5.4.20)$$

Combining (5.4.17), (5.4.18), (5.4.19) and (5.4.20) we obtain the expression for C_{gs} reported in Table 5. 4, where the saturation coefficient α is defined as

$$\alpha = \frac{Q'_{VD}}{Q'_{VS}} = \frac{Q'_{ID} - nC'_{ox}\phi_t + \frac{I_D}{W v_{sat}}}{Q'_{IS} - nC'_{ox}\phi_t + \frac{I_D}{W v_{sat}}}. \quad (5.4.21)$$

The above definition for the saturation coefficient is a generalization of that given in (5.2.11). At the origin of the linear region, ($V_{DS}=0$) $\alpha=1$, and the correction terms that include the saturation velocity vanish as expected.

Following the same procedure as that used for long-channel transistors, but now including the supplementary terms related to velocity saturation, the compact expressions for the (trans)capacitances listed in Table 5.4 are found. In these expressions g_{ms} , g_{md} , and g_{mg} are the source,

drain and gate transconductances, respectively. The use of the generalized saturation coefficient α allows a very compact formulation of the capacitive coefficients.

Fig. 5.19 shows C_{gs} and C_{gd} , normalized to the gate capacitance C_{ox} , versus V_D , calculated according to the expressions of Table 5.4 [13]. These capacitances saturate for lower V_{DS} values when we include velocity saturation, which is consistent with the reduced value of V_{DSsat} obtained when velocity saturation is considered. The plots of other capacitive coefficients are shown in Figs. 20 and 21. In all cases the effect of velocity saturation is to make capacitances vary more gradually as compared to the long channel case.

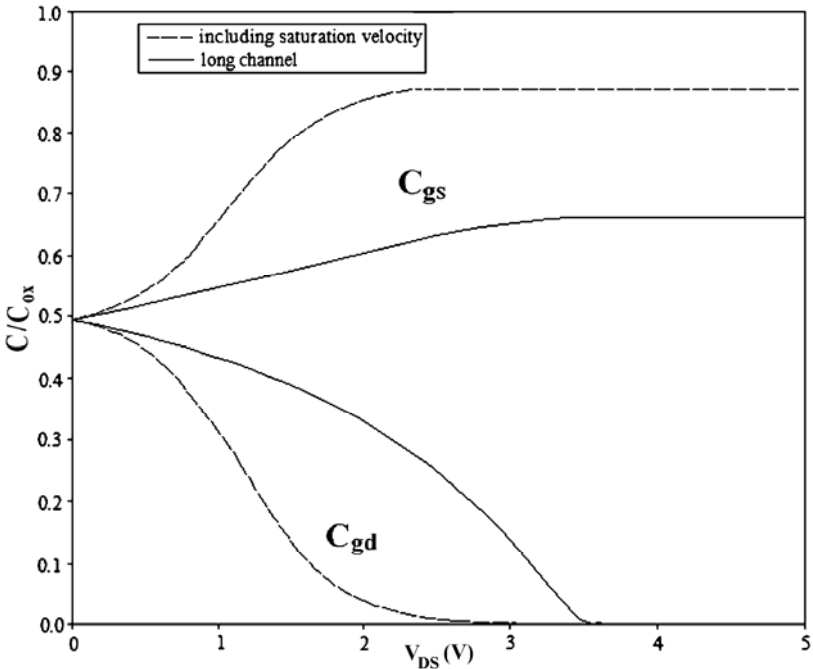


Fig. 5.19. Normalized C_{gs} and C_{gd} calculated (—) neglecting the saturation velocity and (-----) considering saturation velocity ($V_S = V_B = 0V$).

Table 5.4 Intrinsic (trans)capacitances. (After [13].)

Variable	Expression
C_{gs}	$\frac{2}{3}WL_eC'_{ox} \frac{1+2\alpha}{(1+\alpha)^2} \frac{q'_{IS}}{1+q'_{IS}} + \frac{L_e g_{ms}}{3nv_{sat}} \frac{(1-\alpha)^2}{(1+\alpha)^2}$
C_{gd}	$\frac{2}{3}WL_eC'_{ox} \frac{\alpha^2+2\alpha}{(1+\alpha)^2} \frac{q'_{ID}}{1+q'_{ID}} - \frac{L_e g_{md}}{3nv_{sat}} \frac{(1-\alpha)^2}{(1+\alpha)^2}$
$C_{bs(d)}$	$(n-1)C_{gs(d)}$
$C_{gb}=C_{bg}$ (*)	$\frac{n-1}{n} \left(C_{ox} - C_{gso} - C_{gdo} - \frac{L_e g_{mg}}{3v_{sat}} \frac{(1-\alpha)^2}{(1+\alpha)^2} \right)$
C_{ds}	$-\frac{4}{15}nC'_{ox}W \frac{L_e^2}{L} \frac{1+3\alpha+\alpha^2}{(1+\alpha)^3} \frac{q'_{IS}}{1+q'_{IS}} - \frac{1}{30} \frac{g_{ms}L_e^2}{v_{sat}L} \frac{(3\alpha+7)(1-\alpha)^2}{(1+\alpha)^3}$
C_{sd}	$-\frac{4}{15}nC'_{ox}W \frac{L_e^2}{L} \frac{\alpha+3\alpha^2+\alpha^3}{(1+\alpha)^3} \frac{q'_{ID}}{1+q'_{ID}} + \frac{1}{30} \frac{g_{md}L_e^2}{v_{sat}L} \frac{(3+7\alpha)(1-\alpha)^2}{(1+\alpha)^3}$
α	$\frac{Q'_{VD}}{Q'_{VS}} = \frac{Q'_{ID} - nC'_{ox}\phi_t + \frac{I_D}{Wv_{sat}}}{Q'_{IS} - nC'_{ox}\phi_t + \frac{I_D}{Wv_{sat}}}$

(*) C_{gso} and C_{gdo} are the first terms in C_{gs} and C_{gd} , respectively.

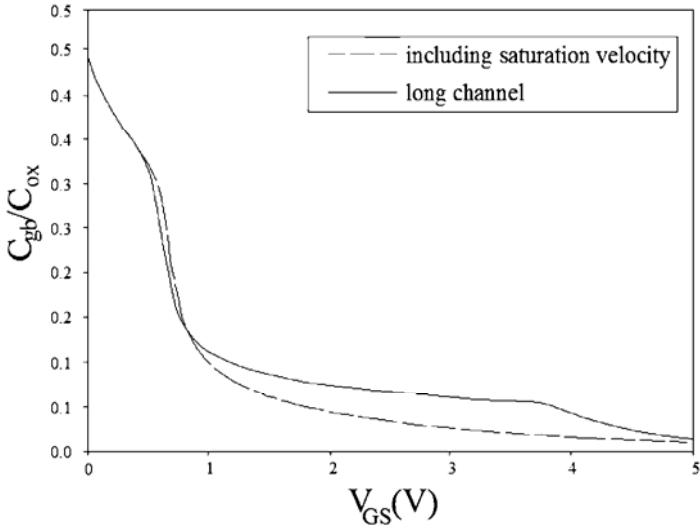


Fig. 5.20 Normalized gate-to-bulk capacitance with and without the effect of velocity saturation. (After [11].)

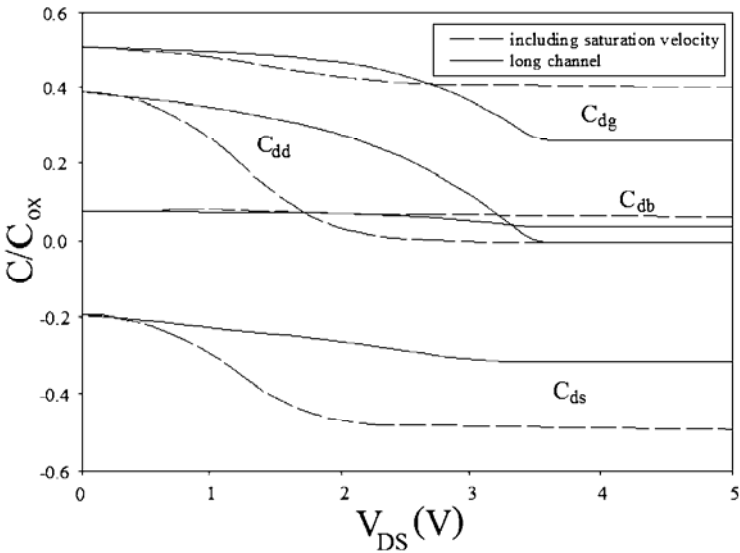


Fig. 5.21 Normalized drain-related capacitances calculated with (----) and (____) without carrier velocity saturation. (After [11].)

5.4.3 Other small-dimension effects on capacitances

Fig. 5.22 shows the dependence of the normalized C_{gd} and C_{gs} on the channel length. Due to DIBL C_{gd} becomes (slightly) negative for the transistor whose channel length is $0.8\mu\text{m}$, the minimum value for this

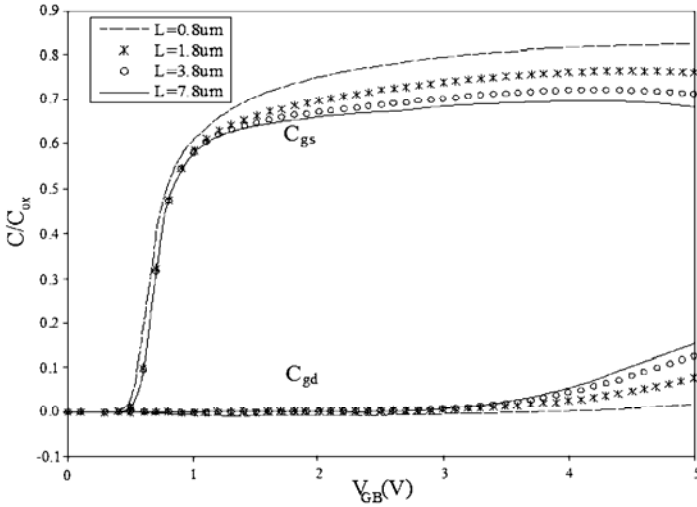


Fig. 5.22 Normalized C_{gs} , C_{gd} vs. V_G for channel length varying from $0.8\mu\text{m}$ to $7.8\mu\text{m}$. $V_D=5\text{V}$ and $V_S=V_B=0\text{V}$. (After [11], [13].)

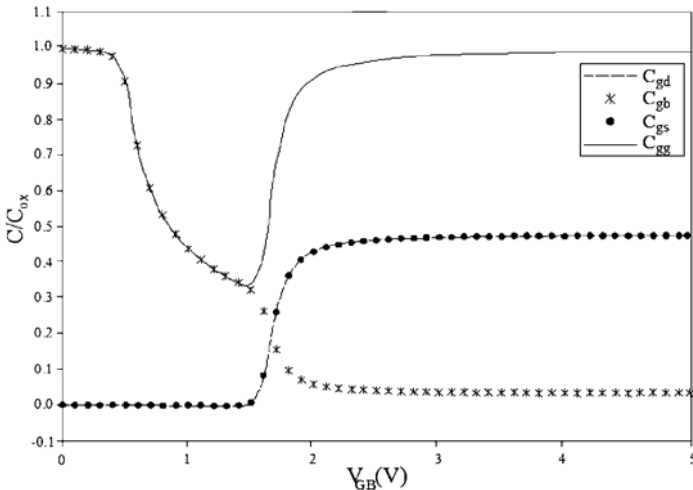


Fig. 5.23 Normalized C_{gs} , C_{gd} , C_{gb} and C_{gg} vs. V_G for $V_D=V_S=V_B=0\text{V}$.

technology. Finally, Fig. 5.23 shows the gate-source and gate-drain capacitances versus V_G for $V_D=V_S=0$. C_{gd} and C_{gs} are equal and the value of the gate capacitance C_{gg} approximates that of the oxide capacitance both in accumulation and strong inversion, as it should. Even if the simplified model of the bulk charge presented here is not valid in accumulation, the analytical implementation in [12] gives a qualitatively correct behavior in this region.

Bibliography

- [1] Soo-Young Oh, Donald E. Ward, and Robert W. Dutton, "Transient analysis of MOS transistors," *IEEE Trans. Electron Devices*, vol. 27, no. 8, pp. 1571-1578, Aug. 1980.
- [2] John E. Meyer, "MOS models and circuit simulation," *RCA Review*, vol. 32, pp. 42-63, Mar. 1971.
- [3] Richard B. Adler, Arthur C. Smith, and Richard L. Longini, *Introduction to Semiconductor Physics*, SEEC Series, vol. 1, John Wiley & Sons, New York, 1964.
- [4] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electron.*, vol. 38, no 11, pp. 1945-1952, Nov. 1995.
- [5] J. J. Paulos and D. A. Antoniadis, "Limitations of quasi-static capacitance models for the MOS transistor," *IEEE Electron Device Lett.*, vol. 4, no. 7, pp. 221-224, July 1983.
- [6] Yannis Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed., McGraw-Hill, Boston, 1999.
- [7] Claudio Turchetti, Guido Masetti and Yannis Tsividis, "On the small signal behavior of the MOS transistor in quasi-static operation," *Solid-State Electron.*, vol. 26, no. 10, pp. 941-949, Oct. 1983.
- [8] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510-1519, Oct. 1998.
- [9] Carlos Galup-Montoro, Márcio Cherem Schneider, and Ana Isabela Araújo. Cunha, "A current-based MOSFET model for integrated circuit design," Chapter 2 of *Low-Voltage/Low-Power Integrated Circuits and Systems*, pp. 7-55, edited by Edgar Sánchez-Sinencio and Andreas G. Andreou, IEEE Press, Piscataway, NJ, 1999.
- [10] Kenneth S. Kundert, *The Designer's Guide to SPICE & SPECTRE*, Kluwer Academic Publishers, Boston, 1995
- [11] Oscar da Costa Gouveia Filho, *Um modelo compacto do transistor MOS para simulação de circuitos*, PhD Thesis, Federal University of Santa Catarina, Florianopolis, Brazil, 1999 (in Portuguese).
- [12] ACM model (1997), [Online]. Available: <http://eel.ufsc.br/lci/acm/index.html>
- [13] Oscar da Costa Gouveia Filho, Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup Montoro, "Advanced compact model for short-channel MOS transistors," *Proc. IEEE CICC 2000*, pp. 209-212.

Problems

- 5.1.** Charge partition: For a long-channel transistor plot the ratio Q_s / Q_i and Q_D / Q_i vs. q'_{IS} for $0.1 < q'_{IS} < 100$, with $q'_{ID} / q'_{IS} = 0, 0.5$, and 1 .
- 5.2.** Verify that any capacitance of the MOSFET can be written as a combination of the nine capacitances in Table 5.3.
- 5.3.** Using the expression in Table 5.3, plot C_{gs} and C_{gd} vs. q'_{IS} for $0.1 < q'_{IS} < 100$ and $q'_{ID} / q'_{IS} = 0, 0.5, 1$.
- 5.4.** Determine the asymptotic values of the capacitances in Table 5.3 for a transistor in strong inversion and saturation.
- 5.5.** Calculate the frequency at which the controlled current source $g_{mg}v_{gb} - C_m dv_{gb} / dt$ in Fig. 5.6 has equal real and imaginary parts. Plot the frequency you calculated vs. the inversion level at the source for a transistor in saturation.
- 5.6.** (a) Using the equivalent circuit in Fig. 5.12, derive the intrinsic cutoff frequency of expression (5.2.79) of a transistor in saturation. (b) Demonstrate that, for $C_{gd} \neq 0$, a slightly modified version of (5.2.79) can be employed to calculate the intrinsic cutoff frequency.
- 5.7.** A short-channel transistor of a CMOS technology ($\mu=200 \text{ cm}^2/\text{V}\cdot\text{s}$, $v_{sat}=10^7 \text{ cm/s}$) has $W/L=2\mu\text{m}/0.1\mu\text{m}$. For $q'_{IS}=20$, plot the curve q'_I / q'_{IS} along the transistor channel for $V_{DS}=0, V_{DSsat}/2, V_{DSsat}$. On top of the curves of the short-channel device, plot q'_I / q'_{IS} of an equivalent long-channel device for the same set of drain-to-source voltages.

This page intentionally left blank

Chapter 6

Mismatch Modeling

Conventional bulk CMOS technology is still (and will be for the next few years) prevalent in the microelectronics industry. According to the International Technology Roadmap for Semiconductors (ITRS 2004 update [1]), bulk MOS transistors will still be used for the 45 nm technology node (gate length around 18 nm), which is expected to be running by 2010. The feasibility of 15 nm conventional MOS transistors in bulk CMOS technology has already been demonstrated [2]. Thus, dopant fluctuation in the depletion region will be the dominant factor for the random variations of next generations bulk CMOS processes. By focusing mainly on the prevailing cause of parameter fluctuations, we intend to provide circuit designers with a mismatch model that is very simple and easy to use for hand-calculations.

6.1 Introduction

Mismatch is the denomination of time-independent variations between identically designed components. The performance of most analog or even digital circuits relies on the concept of matched behavior between identically designed devices [3], [4], [5], [6]. In analog circuits, the spread in the dc characteristics of supposedly matched transistors results in inaccurate or even anomalous circuit behavior [4]. Also, for digital circuits, transistor mismatch leads to propagation delays whose spread can be of the order of several gate delays for deep-submicron technologies [4], [7]. The shrinkage of the MOSFET dimensions and the reduction in the supply voltage make matching limitations even more

important. As predicted by Professor J. Meindl [8], “variations will set the ultimate limits on scaling of MOSFETs.”

Manufacturing variations and the discrete nature of matter and charge result in lot to lot, wafer to wafer, inter-die, and intra-die variations of identically designed devices. Mismatch is the result of either systematic or stochastic (random) effects [9]. Systematic effects are originated by either uncontrollable variation during the fabrication of the integrated circuit or by poor layout. At the origin of systematic mismatch can be equipment-induced non-uniformities such as temperature gradients and photo-mask size differences across the wafer. Systematic effects are important for large distances and can be combated using appropriate layout techniques [9].

Random mismatch refers to local variation in parameters such as doping concentration, oxide thickness, polysilicon granularity, edge irregularity, etc. The first three fluctuations are called areal fluctuations because they scale with device area while the last one is called peripheral fluctuation because it scales with device perimeter [9]. Random mismatch dominates over systematic mismatch for small distances and is generally assumed to display a Gaussian distribution characterized by its standard deviation. Stochastic mismatch requires a model to orient the IC designer regarding sizing and biasing strategies.

The purpose of this chapter is to develop a simple model for random mismatch in bulk CMOS transistors. We will mainly focus on mismatch due to random fluctuations of the dopant concentration, first studied by Keyes [10], which is nowadays recognized as the main cause of mismatch in bulk CMOS transistors.

6.2 Random mismatch in MOS transistors

The most important sources of random mismatch in MOS transistors are edge roughness, oxide granularity, substrate doping, and fixed oxide charges [3], [5], [11]. These error sources contribute to errors in the MOSFET dc electrical parameters such as specific current, threshold voltage, and body effect factor.

To give the reader an insight into how the dopant concentration in advanced technologies affects the electrical performance of a transistor, let us refer to the example presented in [4]. A minimum size transistor in a 0.25 μm CMOS process contains about 1100 dopant atoms in the depletion layer underneath the channel. In a 0.1 μm process, this number is only 200. Assuming a Poisson distribution of impurities in both the 0.25 and 0.1 μm technologies, the spreads in the number of dopant atoms beneath the channel are about 33 and 14, respectively. In both cases, the spread in this number causes a spread in the threshold voltage of about 30 mV, the effect of which increases with new process generation. For operation in weak inversion, a 30 mV deviation in V_T causes around a 2.5-fold deviation in transistor current, which can be catastrophic for a wide number of applications.

It is widely accepted that matching can be modeled by the random variations in geometric, process and/or device parameters, and the effect of these random parameters on the drain current is quantified using the transistor dc model. As pointed out in [12] and more recently in [13], there is a fundamental flaw in the current use of dc models for mismatch that results in inconsistent formulas. In effect, mismatch models implicitly assume that the actual values of the lumped model parameters can be obtained by integration of the position-dependent parameters distributed over the channel region of the device, *e.g.*, for the threshold voltage V_T

$$V_T = \frac{1}{WL} \iint_{\text{channel-area}} V_T(x, y) dx dy \quad (6.2.1)$$

where W and L are the width and length of the transistor.

As analyzed in [12], the use of lumped parameters for the series or parallel association of transistors leads to an inconsistent model of mismatch owing to the nonlinear nature of MOSFETs. The conventional approach [5] to modeling mismatch takes into account the dopant fluctuations over the entire channel but here we consider explicitly the effects of local fluctuations. We integrate the contribution of the local fluctuations along the channel considering the main MOSFET nonlinearities. Fortunately, the formalism needed to include local

fluctuations is already available in flicker or 1/f noise modeling [14], namely carrier number fluctuation theory.

6.3 Consistent model for drain current fluctuation

For a given set of bias voltages, the fluctuations of the drain current around its nominal value result from the sum of all the contributions from local fluctuations along the channel, whatever their origin. To calculate the effect of these fluctuations, we split the transistor into 3 series elements as shown in Fig. 6.1(a): an upper transistor, a lower transistor, and a small channel element of length Δy and area $\Delta A = W\Delta y$. In Fig. 6.1(a), y is the distance from the channel element to the source.

The local current fluctuation ($i_{\Delta A}$) is assumed to be a zero-mean stationary random process on the variable y . Small-signal analysis allows one to calculate the effect of $i_{\Delta A}$ on the drain current deviation (ΔI_d), as shown in Fig. 6.1(b). Using (3.7.25), it follows that

$$g_u = -\mu \frac{W}{L-y} Q'_l \quad (6.3.1)$$

and

$$g_l = -\mu \frac{W}{y} Q'_l, \quad (6.3.2)$$

the current division between the channel element and the equivalent small-signal resistance of the rest of the channel gives $\Delta I_d = (\Delta y/L)i_{\Delta A}$. This very simple result for the current division, proportional to a geometric ratio, is a consequence of the quasi-Fermi potential formulation for the drain current, *i.e.*, the conductance of the channel element and the transconductances of the upper and lower transistors are proportional to the local charge density [15], [16], [17]. Thus, the mean square of the total drain current fluctuation is

$$\overline{\Delta I_d^2} = \sum_{\text{channel-length}} \overline{(\Delta I_d)^2} = \lim_{\Delta y \rightarrow 0} \sum \left(\frac{\Delta y}{L} i_{\Delta A} \right)^2 = \frac{1}{L^2} \int_0^L \Delta y \overline{i_{\Delta A}^2} dy \quad (6.3.3)$$

since local current fluctuations along the channel are uncorrelated. Local current fluctuations arise from three independent physical origins, namely fluctuations of channel doping, surface state density, and gate oxide thickness [18]. Note that since $i_{\Delta A}$ is related to local fluctuations in the area $W\Delta y$, its variance is proportional to $I/W\Delta y$. As in reference [18], we have assumed that fluctuation of channel doping is the main factor that determines local current fluctuations.

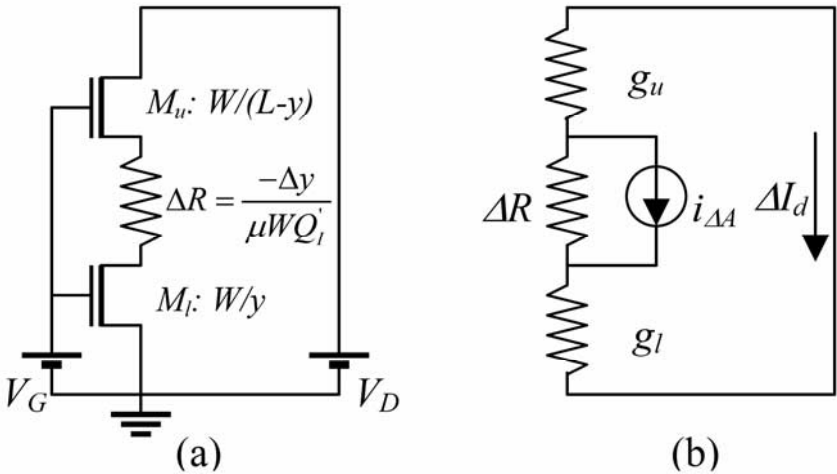


Fig. 6.1 Splitting of a transistor into three series elements: (a) transistor equivalent circuit and (b) small-signal equivalent circuit.

6.4 Number fluctuation mismatch model

The local current fluctuation that results from local fluctuations in the inversion charge density is calculated from

$$I_D = -\mu \frac{W}{dy} Q'_l(y) dV_c \quad (6.4.1)$$

as in [19], [20], [21], yielding

$$i_{\Delta A} = I_D \frac{\Delta Q'_l}{Q'_l} \quad (6.4.2)$$

where $\Delta Q'_I$ is the fluctuation of the inversion charge density in a channel element of area ΔA . For the sake of simplicity we will consider only the fluctuations in the number of carriers, but the analysis can also be extended to include mobility fluctuation, as carried out in [19] for $1/f$ noise.

As in [18], we assume that a fluctuation in the number of impurities is solely responsible for a fluctuation in the number of carriers. To derive the fluctuation of the inversion charge density, we have used the capacitive MOS model of Fig. 6.2.

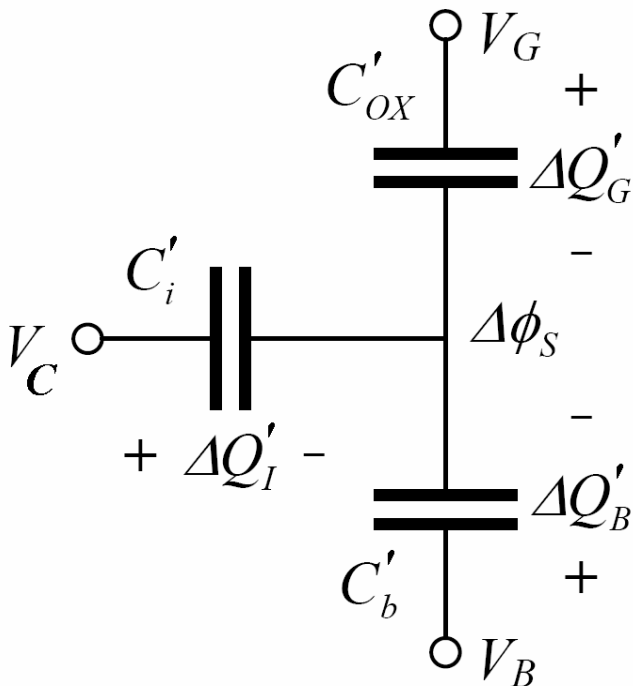


Fig. 6.2 Capacitive model of the MOSFET channel for matching analysis. Terminal voltages are constant.

Charge conservation in the structure of Fig. 6.2 requires that

$$\Delta Q'_B(y) + \Delta Q'_G(y) + \Delta Q'_I(y) = 0 \quad (6.4.3)$$

where $\Delta Q'_G = -C'_{ox} \Delta \phi_s$ and $\Delta Q'_I = -C'_i \Delta \phi_s$. The variation in the depletion charge is the sum of two components, the first equal to $-C'_B \Delta \phi_s$, associated with the fluctuation in the surface potential, and the second, designated by $\Delta Q'_{IMP}$, associated with the fluctuation in the number of ionized impurities. Therefore, the variation in the depletion charge is $\Delta Q'_B = -C'_B \Delta \phi_s + \Delta Q'_{IMP}$.

The definitions of capacitances along with (6.4.3) result in

$$\Delta Q'_I = -\frac{C'_i}{C'_i + C'_B + C'_{ox}} \Delta Q'_{IMP}. \quad (6.4.4)$$

Recalling [15], [22] that $C'_i \cong -Q'_I / \phi_t$ and $C'_B = (n-1)C'_{ox}$ under any bias condition, expression (6.4.4) can be rewritten as

$$\Delta Q'_I = -\frac{Q'_I}{Q'_I - nC'_{ox}\phi_t} \Delta Q'_{IMP}. \quad (6.4.5)$$

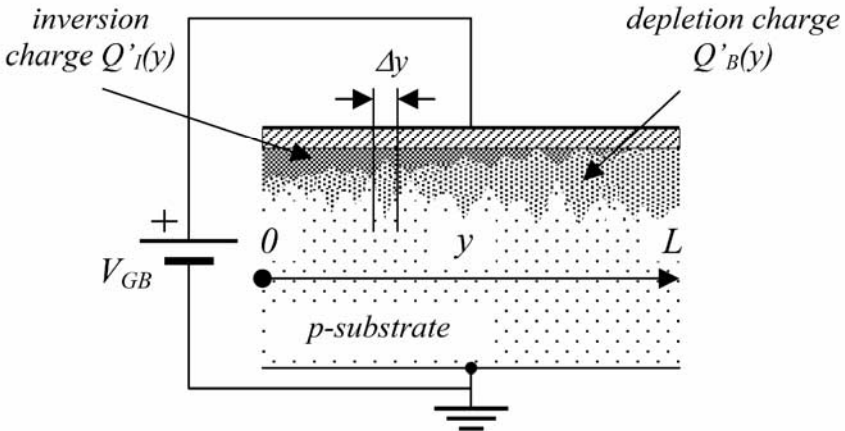


Fig. 6.3 Cross-section of an MOS transistor showing the (greatly exaggerated) fluctuations in both inversion and depletion charge densities due to local dopant fluctuations.

The local fluctuation in the number of impurities in an elementary slab of the depletion layer is calculated assuming it to be a random

variable with Poisson distribution [11], [18]. Fig. 6.3 shows the (greatly exaggerated) fluctuations in the inversion charge density Q'_i due to local dopant fluctuations. Note that both the depletion and inversion charge densities Q'_b and Q'_i change as a result of the variation in the number of impurity atoms along the y-axis.

The average number ($\overline{n_a}$) of dopants in the elementary volume of a depletion layer of length Δy [18] and depth Δx is

$$\overline{n_a} = N_a \Delta x W \Delta y. \quad (6.4.6)$$

In (6.4.6), N_a is the net concentration of dopants (acceptors and donors) in the elementary volume of the depletion layer.

The square of the standard deviation of a random variable with Poisson distribution is equal to its average value; thus,

$$\sigma^2(n_a) = N_a \Delta y W \Delta x. \quad (6.4.7)$$

Now, to calculate the standard deviation of $\Delta Q'_{IMP}$ we proceed as in [23], [24], which assume the individual contributions of the local deviations in the number of impurities to $\Delta Q'_{IMP}$ to be uncorrelated, thus yielding

$$\sigma^2(\Delta Q'_{IMP}) = \frac{q^2}{W \Delta y} \int_0^{x_d} N_a \left(1 - \frac{x}{x_d}\right)^2 dx, \quad (6.4.8)$$

where q is the electron charge, x is the distance to the semiconductor-oxide interface and x_d is the depletion depth. From now on, the integral in (6.4.8) will be written as

$$N_{oi} = \int_0^{x_d} N_a \left(1 - \frac{x}{x_d}\right)^2 dx, \quad (6.4.9)$$

which is an “equivalent” impurity density (number of impurities per unit area) in the depletion layer for the calculation of mismatch. For the simple case of uniform doping, $N_{oi} = N_a x_d / 3$.

Using (6.4.2), (6.4.5), and (6.4.8), $(i_{AA})^2$ can be calculated and, inserting the resulting value into (6.3.3), we obtain the expression for $\overline{\Delta I_D^2}$. Now, using (5.2.3), rewritten here

$$dy = -\frac{\mu W}{nC'_{ox}I_D}(Q'_I - nC'_{ox}\phi_t)dQ'_I, \quad (6.4.10)$$

the integration over the channel length in (6.3.3) is changed to the integration over the channel charge density given by

$$\sigma_{I_D}^2 = \overline{\Delta I_D^2} = \frac{q^2 \mu I_D}{nC'_{ox}L^2} \int_{Q'_{IS}}^{Q'_{ID}} \frac{N_{oi} dQ'_I}{nC'_{ox}\phi_t - Q'_I}. \quad (6.4.11)$$

Expression (6.4.11) allows the computation of current mismatch in terms of both the doping concentration along the depletion region and bias, in this case represented by the inversion charge densities at source and drain. The main difficulty in calculating the integral in (6.4.11) arises from the non-uniform doping profile and from the variation in the depletion depth along the channel. For a constant doping profile, we have verified that the introduction of a variable depletion depth along the channel is generally not relevant, except for very high inversion levels. In saturation, the impact of a variable depletion depth on mismatch is, in a typical 0.35 μm CMOS technology, less than 6% for an inversion level of 1000 as compared with the result obtained assuming a uniform depletion depth. For lower inversion levels, the error becomes smaller.

In order to simplify the notation and derive simple expressions for mismatch, let us assume N_{oi} to be constant. Thus, the integration of (6.4.11) from source to drain results in

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{q^2 N_{oi} \mu}{L^2 n C'_{ox} I_D} \ln \left(\frac{n C'_{ox} \phi_t - Q'_{IS}}{n C'_{ox} \phi_t - Q'_{ID}} \right). \quad (6.4.12)$$

The result in (6.4.12) is essentially the same as that derived for flicker noise in MOS transistors in [14]. This is because mismatch is similar to a “dc noise” and the physical origin of both matching and 1/f noise is related to fluctuations in either fixed charges or localized states along the channel.

Once we have determined the effect of the doping concentration on mismatch, we now proceed to calculate the influence of the variation of the remaining factors, namely oxide thickness, body effect factor, mobility, and dimensions on mismatch. To do that, we are going to

assume that the influence of these factors on mismatch can be simplified by assuming them to affect the transistor specific current only.

6.5 Specific current mismatch

Let us now review the long-channel transistor equations for a brief discussion on the influence of the specific current on transistor mismatch. For convenience, we reproduce below the long-channel equation

$$I_D = I_S(i_f - i_r). \quad (6.5.1)$$

The specific current I_S is given by

$$I_S = \mu C'_{ox} n \frac{\phi_t^2}{2} \frac{W}{L}. \quad (6.5.2)$$

Assuming the temperature to be the same for the devices under analysis, one can write the relative deviation of the specific current as

$$\frac{\Delta I_S}{I_S} = \frac{\Delta \mu}{\mu} - \frac{\Delta t_{ox}}{t_{ox}} + \frac{\Delta n}{n} + \frac{\Delta W}{W} - \frac{\Delta L}{L}. \quad (6.5.3)$$

In a simple model, let us assume the first three terms in (6.5.3) to be independent of bias and the last two terms to represent peripheral fluctuations. We now group the first three terms and rewrite (6.5.3) as

$$\frac{\Delta I_S}{I_S} = \frac{\Delta I_{SQ}}{I_{SQ}} + \frac{\Delta W}{W} - \frac{\Delta L}{L}, \quad (6.5.4)$$

where I_{SQ} is the sheet specific current. We now suppose the errors in (6.5.4) to be uncorrelated. We can then write the standard deviation of the specific current as

$$\frac{\sigma^2(I_S)}{I_S^2} = \frac{\sigma^2(I_{SQ})}{I_{SQ}^2} + \frac{\sigma^2(W)}{W^2} + \frac{\sigma^2(L)}{L^2}. \quad (6.5.5)$$

The result in (6.5.5) is similar to that presented in [3], [5], [11]. We now suppose the standard deviations in (6.5.5) to follow the simple rules

$$\frac{\sigma^2(I_{SQ})}{I_{SQ}^2} = \frac{A_{ISQ}^2}{WL}; \sigma^2(W) = \frac{A_W^2}{L}; \sigma^2(L) = \frac{A_L^2}{W}, \quad (6.5.6)$$

which results in the following expression for the standard deviation of I_S

$$\frac{\sigma^2(I_S)}{I_S^2} = \frac{A_{ISQ}^2}{WL} + \frac{A_W^2}{W^2L} + \frac{A_L^2}{WL^2}. \quad (6.5.7)$$

For long and wide channels, the result in (6.5.7) can be further simplified to

$$\frac{\sigma^2(I_S)}{I_S^2} \cong \frac{A_{ISQ}^2}{WL}, \quad (6.5.8)$$

which is a result similar to the one shown in [5]. Note that, in our model, the relative deviation in the specific current is equivalent to the relative deviation in β given in [5].

6.6 Mismatch model in terms of inversion level

An alternative expression for (6.4.12) which may be of great use to circuit designers is obtained if the charge density, at source and drain, is written in terms of the forward and reverse currents [17], as given in Chapter 4 and repeated here for convenience

$$q'_{IS(D)} = -\frac{Q'_{IS(D)}}{nC'_{ox}\phi_t} = \sqrt{1+i_{f(r)}} - 1. \quad (6.6.1)$$

Using (6.6.1) in (6.4.12), the following expression results for the current mismatch due to fluctuations in the number of dopant atoms

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{N_{oi}}{WLN^{*2}} \frac{1}{i_f - i_r} \ln\left(\frac{1+i_f}{1+i_r}\right), \quad (6.6.2)$$

where we define N^* as in [14], [19], [25]

$$N^* = \frac{-Q'_{IP}}{q} = \frac{nC'_{ox}\phi_t}{q}, \quad (6.6.3)$$

Expression (6.6.2) indicates that the ratio of mismatch power to dc power is inversely proportional to the gate area, as is well-known. The ratio of mismatch power to dc power is proportional to t_{ox}^2 and to $N_a^{1/2}$ for a constant impurity profile. Finally, the substrate voltage also affects the factor N_{oi} through modulation of the depletion depth. For fixed i_f and i_r , reverse substrate-to-source voltages increase the depletion depth and, consequently, N_{oi} . As a result, matching worsens for a reverse biased bulk-to-source junction.

Expression (6.6.2) can be simplified under specific conditions. From weak to strong inversion in the linear region, $i_f \cong i_r$ and (6.6.2) reduces to

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{N_{oi}}{WLN^{*2}} \frac{1}{1+i_f}. \quad (6.6.4)$$

Expression (6.6.4) can also be written as

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{N_{oi}}{WLN^{*2}} \left(n\phi_t \frac{g_m}{I_D} \right)^2. \quad (6.6.5)$$

In weak inversion, $i_f \ll 1$; thus, the first order series expansion of (6.6.2) leads to

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{N_{oi}}{WLN^{*2}}, \quad (6.6.6)$$

for either saturation or nonsaturation. In saturation, $i_r \rightarrow 0$; thus, expression (6.6.2) can be written as

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{N_{oi}}{WLN^{*2}} \frac{\ln(1+i_f)}{i_f}. \quad (6.6.7)$$

Expression (6.6.7) can also be written as

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{N_{oi}}{WLN^{*2}} \frac{\ln(1+I_D/I_S)}{I_D/I_S}. \quad (6.6.8)$$

For the sake of completeness, one can include the random errors due to the specific sheet current $I_{SQ} = \mu C'_{ox} n\phi_t^2 / 2$ as in expression (6.5.8), which results in a modification of expression (6.6.2), yielding

$$\frac{\sigma_{I_D}^2}{I_D^2} = \frac{1}{WL} \left[\frac{N_{oi}}{N^{*2}} \frac{1}{i_f - i_r} \ln \left(\frac{1+i_f}{1+i_r} \right) + A_{ISQ}^2 \right]. \quad (6.6.9)$$

In (6.6.9) A_{ISQ} is a mismatch factor that, in a first approximation, is a constant factor that accounts for variations in mobility, gate oxide thickness, and slope factor. A more elaborate model would consider A_{ISQ} as a (slightly) bias-dependent term.

The current mismatch in expression (6.6.9) can be referred to the gate. To derive the voltage mismatch between current biased transistors, we note that $dI_D = g_{mg} dV_G$; therefore, the voltage mismatch between current biased transistors is

$$\sigma_{V_G}^2 = \frac{I_D^2}{g_{mg}^2} \frac{\sigma_{I_D}^2}{I_D^2}. \quad (6.6.10)$$

Expression (6.6.10) is useful for finding the voltage mismatch in a differential pair, for example.

6.7 Experimental results and discussion [25], [26]

Intra-die current mismatch of a set of NMOS and PMOS transistors was measured on a test circuit fabricated with the TSMC 0.35 μ m 3.3V CMOS n-well process (from MOSIS – www.mosis.org), in which the gate oxide thickness is 78 Å. In the test circuit, transistors are arranged in arrays of 20 identical devices terminated by dummy ones to ensure the same surroundings for all the transistors. Matched transistors have the same orientation. Transistor dimensions ($W \times L$) in the different arrays are 12 μ m x 8 μ m (*large*), 3 μ m x 2 μ m (*medium*), 0.75 μ m x 8 μ m (*narrow* - minimum width), 12 μ m x 0.5 μ m (*short* - minimum length) and 0.75 μ m x 0.5 μ m (*small* - minimum size). Wide metal connections and multiple contact windows were employed in the layout to lower ohmic drops. All of the ten packaged dies out of forty that were characterized showed similar mismatch behavior. Fig. 6.4 shows a microphotograph of the test chip.

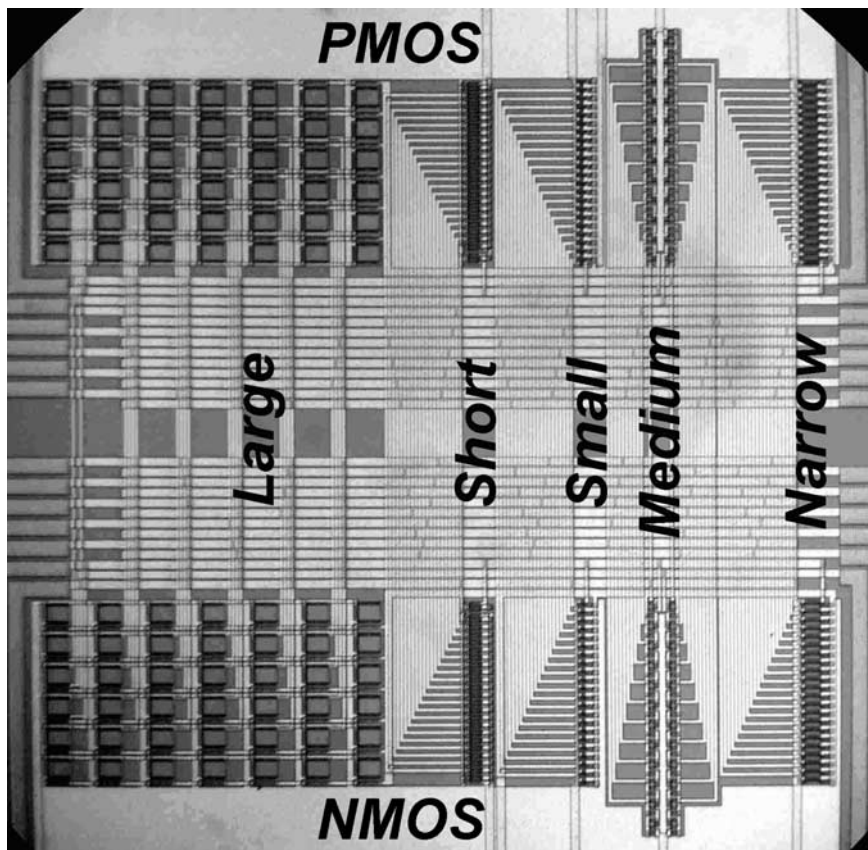


Fig. 6.4 Test chip microphotograph, showing PMOS and NMOS transistor arrays fabricated in a $0.35\ \mu\text{m}$ technology. Transistor dimensions ($W \times L$) in the different arrays are $12\ \mu\text{m} \times 8\ \mu\text{m}$ (*large*), $3\ \mu\text{m} \times 2\ \mu\text{m}$ (*medium*), $0.75\ \mu\text{m} \times 8\ \mu\text{m}$ (*narrow* - minimum width), $12\ \mu\text{m} \times 0.5\ \mu\text{m}$ (*short* - minimum length) and $0.75\ \mu\text{m} \times 0.5\ \mu\text{m}$ (*small* - minimum size).

Fig. 6.5 presents the mismatch power normalized to the dc power for drain-to-source voltage ranging from 10mV (linear region) to 2V (saturation) for the *medium* size NMOS devices. Mismatch was measured for six different inversion levels (0.01, 0.1, 1, 10, 100, and 1000). The bulk terminal was kept at zero volts. Simulated (model) curves were determined from expression (6.6.2), with i_r calculated through expression (3.7.20).

In weak inversion ($i_f = 0.01$ and 0.1), mismatch is almost constant from the linear to saturation region and independent of the inversion level, as predicted by (6.6.6). Measured and simulated curves for weak inversion are almost coincident, being hardly distinguishable. From moderate ($i_f = 1$ and 10) to strong ($i_f = 100$) inversion, both the simulated and measured curves present similar behavior, increasing from the linear to saturation region, where a plateau is reached. Differences between measured and simulated curves at saturation may be associated with statistical spatial-nonuniformity concentration of dopant atoms.

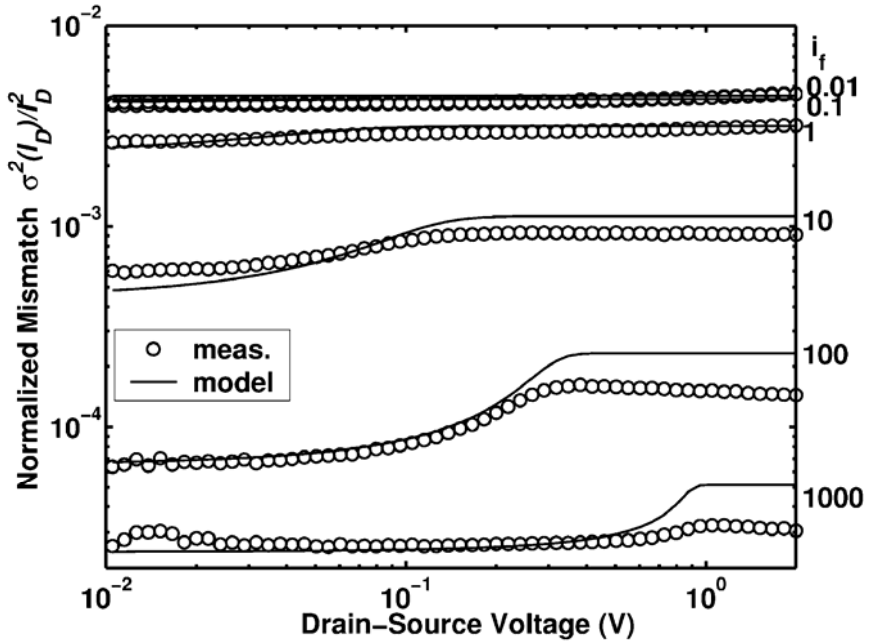


Fig. 6.5 Normalized current mismatch power for the *medium-sized* ($W= 3 \mu\text{m}$, $L= 2 \mu\text{m}$) NMOS transistor array.

Parameter N_{oi} was estimated from measurements in weak inversion, using equation (6.6.6), with effective transistor width and length. N^* was calculated based on parameters provided by MOSIS. The same value of N_{oi} , $1.8 \times 10^{12} \text{ cm}^{-2}$ for the NMOS devices, and $7 \times 10^{12} \text{ cm}^{-2}$ for the PMOS devices, was obtained for both the *large* and *medium* transistors. It

should be noted that N_{oi} includes both the acceptor and donor impurities [27]. As a consequence, N_{oi} is usually higher than the product of the net ion concentration and the depletion layer depth.

At an inversion level $i_f = 1000$, mismatch calculated using (6.6.2) deviates considerably from the experimental results. In fact, when the inversion level is high, the inversion charge layer (channel) provides a “shield” for the gate-bulk electric field. This leads to the influence of the random dopant placement on mismatch being small, and the effects of other components of mismatch such as variation in gate oxide thickness, mobility, and slope factor then play an important role. Therefore, the more complete expression (6.6.9) must be used to account for mismatch factors other than doping fluctuations. For high inversion levels, mismatch flattens out at a minimum value determined by A_{ISQ} , a result that was corroborated by the experimental data.

Parameter A_{ISQ} was estimated from measurements in strong inversion and the linear region, using (6.6.9). A_{ISQ} values of the order of 0.89 %/ μm and 0.71 %/ μm resulted for NMOS and PMOS devices, respectively, for both the *large* and the *medium* devices.

The simulated curves presented in Fig. 6.5 and Fig. 6.6 are based on the values extracted for both N_{oi} and A_{ISQ} for either NMOS or PMOS transistors. Fig. 6.6 shows the measured and simulated dependence of current matching on inversion level (or bias current I_B) for the linear and saturation regions, for three sizes of NMOS and PMOS transistors.

From Fig. 6.6, one can see that larger transistors follow the “area rule”, as shown in our model. We also used the same A_{ISQ} value for modeling the matching of both the *large* and *medium* devices. The *small* transistors do not follow this rule, presenting a mismatch 55% lower (NMOS) and 80% higher (PMOS) than the model estimates using the same N_{oi} as that for the large and medium-sized transistors. However, values of N_{oi} for the *small* transistors different from those measured for the *large* transistors were chosen in order to obtain better fitting of the curves. For the dies we characterized, *small* transistors presented an unpredictable N_{oi} . In fact, electrical characteristics of short-channel devices are very sensitive to fluctuations due to a greater dependence on edge effects. This high sensitivity of short-channel devices is one of the main reasons for the difficulties found in modeling mismatch, mainly in today’s very complex submicron technologies. Also, for minimum length

devices, drain and source doped regions are very close to each other, affecting strongly the shape of the depletion layer below the channel.

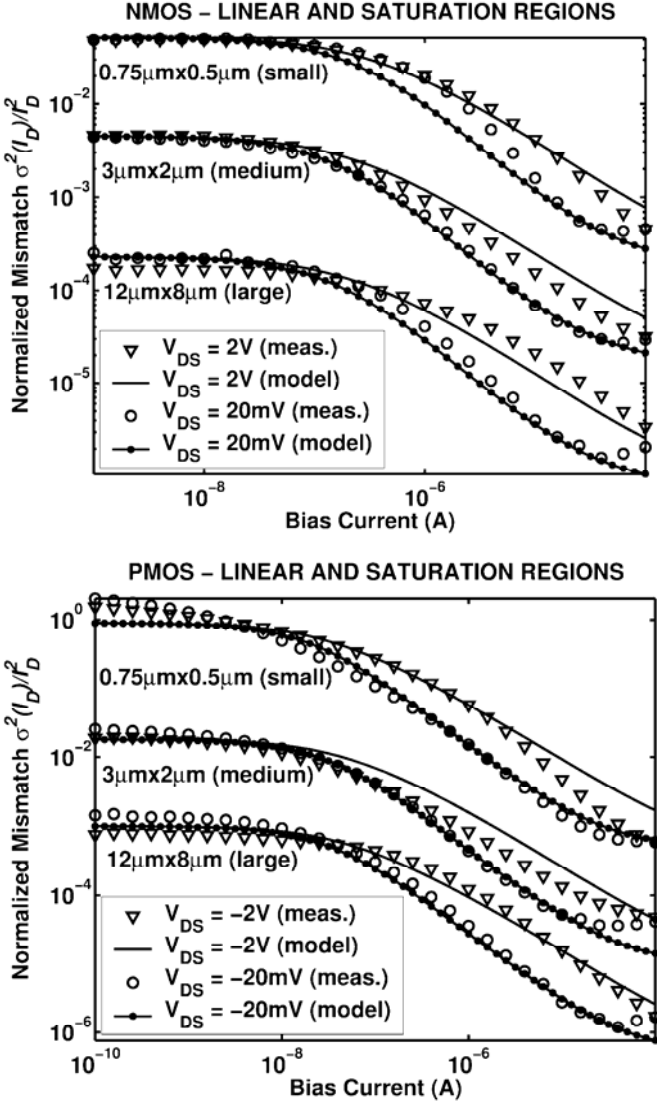


Fig. 6.6 Dependence of current matching on inversion level (bias current I_B) in linear and saturation regions ($|V_{DS}|=20\text{mV}$ and $|V_{DS}|=2\text{V}$, respectively) for the *large*-, *medium*-, and *small*-sized NMOS and PMOS transistor arrays.

The curves presented in Fig. 6.6 have a similar behavior to that seen in $1/f$ noise characterization [14], showing that both phenomena, mismatch and $1/f$ noise, arise from the same mechanism, although the first is related to random doping density fluctuation and the second to random trapping-detrapping of carriers in the channel oxide-substrate interface.

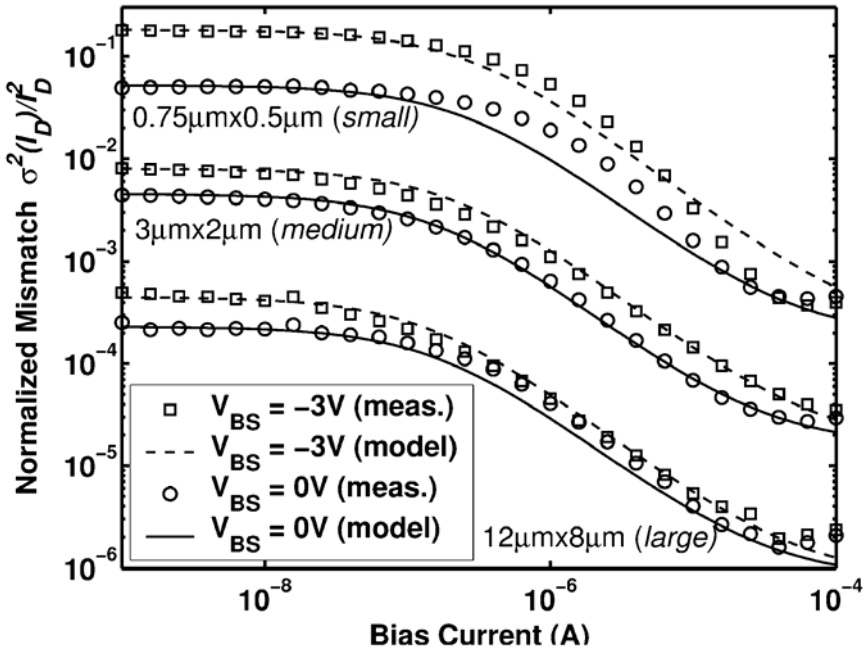


Fig. 6.7 Dependence of current matching on inversion level in the linear region for the *large*, *medium* and *small* NMOS transistor arrays at two bulk-source voltages.

Fig. 6.7 shows the measured and simulated dependence of current matching on inversion level for the linear region, for three sizes of NMOS transistors, at two bulk bias voltages (V_{BS}). Larger transistors follow the “area rule”, as our mismatch model predicts. For a particular bulk bias, we used the same N_{oi} for modeling the matching of both the *large* and *medium* transistors. We also used the same A_{ISQ} value for modeling the matching of both the *large* and *medium* devices under any

bulk bias. The *small* transistors do not follow the “area rule”, presenting a mismatch 55% lower than the model estimates (at zero volt bulk bias) using the same N_{oi} . At $V_{BS} = -3V$, the mismatch measured for the *small* transistors is in good agreement with the value estimated by our model. However, values of N_{oi} for the *small* transistors different from those measured for the *large* transistors were chosen in order to obtain better fitting of the curves. For the dies we characterized, *small* transistors presented an unpredictable N_{oi} , as previously observed in references [7], [28]. As experimental data have demonstrated, the model we developed for mismatch can also be used for short-channel transistors, even though fitting of both N_{oi} and A_{ISQ} is required. A good approach for modeling mismatch in short-channel transistors would be to define a range of “maximum-minimum” values for N_{oi} [29]. In a conservative design, the maximum value of N_{oi} would be chosen to predict the worst-case mismatch.

From Fig. 6.7, one can see that current mismatch increases for reverse bulk bias. The reason for this behavior is that, compared to zero bulk bias, a reverse bulk voltage deepens the depletion layer, resulting in a higher N_{oi} [30], [31], [32], [33]. Owing to the difficulty in developing a model for the modulation of N_{oi} with bulk bias, we decided to choose values of N_{oi} that fit the measured data for each bulk bias condition.

For more details about the experimental results we have reported here, the reader is referred to [25].

Our measurements span a wide range of six decades of current, going from very weak to very strong inversion, and fit well the mismatch over the whole of this current range. Also, we have demonstrated the accuracy of our mismatch model for both the linear and saturation regions. Better accuracy could be attained using more fitting parameters, but we preferred to keep our mismatch model simple, making it a useful hand-design tool that requires only two parameters (N_{oi} and A_{ISQ}) to interface technology with designers.

Even though the model we have derived is different from the popular Pelgrom’s [5] model, we can associate the parameters of the model presented in this text with those presented in [5], as given below

$$A_{\beta}^2 = A_{ISQ}^2. \quad (6.7.1)$$

$$A_{VT}^2 = \frac{q^2}{C_{ox}^2} N_{oi}. \quad (6.7.2)$$

At this point, we emphasize that the mismatch model presented here is continuous in all operating regions and consistent with the distributed nature of the MOS transistor. As a result, the mismatch model derived here is fully consistent with the series-parallel association of transistors.

Bibliography

- [1] International Technology Roadmap for Semiconductors (2004 update). [Online]. Available: <http://public.itrs.net/>
- [2] Bin Yu, Haihong Wang, Amol Joshi, Qi Xiang, Effiong Ibok, and Ming-Ren Lin, "15 nm gate length planar CMOS transistor," *IEDM Tech. Dig.*, 2001, pp. 937-939.
- [3] Jyn-Bang Shyu, Gabor C. Temes, and François Krummenacher, "Random error effects in matched MOS capacitors and current sources," *IEEE J. Solid-State Circuits*, vol. 19, no. 6, pp. 948-955, Dec. 1984.
- [4] Harry Veendrick, *Deep-Submicron CMOS ICs*, 2nd ed., Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [5] Marcel J. M. Pelgrom, Aad C. J. Duinmaijer, and Anton P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433-1440, Oct. 1989.
- [6] Francesco Forti and Michael E. Wright, "Measurements of MOS current mismatch in the weak inversion region," *IEEE J. Solid-State Circuits*, vol. 29, no. 2, pp. 138-142, Feb. 1994.
- [7] Marcel J. M. Pelgrom, Hans P. Tuinhout, and Maarten Vertregt, "Transistor matching in analog CMOS applications," *IEDM Tech. Dig.*, 1998, pp. 915-918.
- [8] Ron Wilson, "The dirty little secret: Engineers at design forum vexed by rise in process variations at the die level," *EE Times*, Mar. 25, 2002, p. 1.
- [9] Alan Hastings, *The Art of Analog Layout*, Prentice Hall, Upper Saddle River, NJ, 2001.
- [10] Robert W. Keyes, "Effect of randomness in the distribution of impurity ions on FET thresholds in integrated electronics," *IEEE J. Solid-State Circuits*, vol. 10, no. 4, pp. 245-247, Aug. 1975.
- [11] Kadaba R. Lakshmikumar, Robert A. Hadaway, and Miles A. Copeland, "Characterization and modeling of mismatch in MOS transistors for precision analog design," *IEEE J. Solid-State Circuits*, vol. 21, no. 6, pp. 1057-1066, Dec. 1986.
- [12] Mao-Feng Lan and Randall Geiger, "Impact of model errors on predicting performance of matching-critical circuits," *Proc. 43rd IEEE MWSCAS*, 2000, pp. 1324-1328.

- [13] Hongning Yang, Véronique Macary, John L. Huber, Won-Gi Min, Bob Baird, and Jiangkai Zuo, "Current mismatch due to local dopant fluctuations in MOSFET channel," *IEEE Trans. Electron Devices*, vol. 50, no. 11, pp. 2248-2254, Nov. 2003.
- [14] Alfredo Arnaud and Carlos Galup-Montoro, "A compact model for flicker noise in MOS transistors for analog circuit design," *IEEE Trans. Electron Devices*, vol. 50, no. 8, pp. 1815-1818, Aug. 2003.
- [15] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510-1519, Oct. 1998.
- [16] Yannis Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed., McGraw-Hill, Boston, 1999.
- [17] Carlos Galup-Montoro, Márcio Cherem Schneider, and Ana Isabela Araújo Cunha, "A current-based MOSFET model for integrated circuit design," Chapter 2 of *Low-Voltage/Low-Power Integrated Circuits and Systems*, pp. 7-55, edited by Edgar Sánchez-Sinencio and Andreas G. Andreou, IEEE Press, Piscataway, NJ, 1999.
- [18] Tomohisa Mizuno, Jun-ichi Okamura, and Akira Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 2216-2221, Nov. 1994.
- [19] Kwok K. Hung, Ping K. Ko, Chenming Hu, and Yiu C. Cheng, "A physics-based MOSFET noise model for circuit simulators," *IEEE Trans. Electron Devices*, vol. 37, no. 5, pp. 1323-1333, May 1990.
- [20] G. Reimbold, "Modified 1/f trapping noise theory and experiments in MOS transistors biased from weak to strong inversion – influence of interface states," *IEEE Trans. Electron Devices*, vol. 31, no. 9, pp. 1190-1198, Sep. 1984.
- [21] F. Berz, "Theory of low frequency noise in Si MOST's," *Solid-State Electron.*, vol. 13, no. 5, pp. 631-647, May 1970.
- [22] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "Derivation of the unified charge control model and parameter extraction procedure," *Solid-State Electron.*, vol. 43, no. 3, pp. 481-485, Mar. 1999.
- [23] P. A. Stolk and D. B. M. Klaassen, "The effect of statistical dopant fluctuations on MOS device performance," *IEDM Tech. Dig.*, 1996, pp. 627-630.
- [24] Kiyoshi Takeuchi, Toru Tatsumi, and Akio Furukawa, "Channel engineering for the reduction of random-dopant-placement-induced threshold voltage fluctuation," *IEDM Tech. Dig.*, 1997, pp. 841-844.
- [25] Carlos Galup-Montoro, Márcio C. Schneider, Hamilton Klimach, and Alfredo Arnaud, "A compact model of MOSFET mismatch for circuit design," *IEEE J. Solid-State Circuits*, vol. 40, no. 8, pp. 1649-1657, Aug. 2005.
- [26] Hamilton Klimach, Carlos Galup-Montoro, Márcio C. Schneider and Alfredo Arnaud, "MOSFET mismatch modeling: A new approach," *IEEE Design and Test of Computers*, vol. 23, no. 1, pp. 20-29, Jan.-Feb. 2006.
- [27] Marcel J. M. Pelgrom, "Low-power CMOS data conversion," Chap. 14 in *Low-Voltage/Low-Power Integrated Circuits and Systems*, pp. 432-484, edited by Edgar Sánchez-Sinencio and Andreas G. Andreou, IEEE Press, Piscataway, NJ, 1999.

- [28] M. Steyaert, J. Bastos, R. Roovers, P. Kinget, W. Sansen, B. Graindourze, A. Pergoot, E. Janssens., "Threshold voltage mismatch in short-channel MOS transistors," *Electron. Lett.*, vol. 30, no. 18, pp. 1546-1548, Sep. 1994.
- [29] José Pineda de Gyvez and Hans P. Tuinhout, "Threshold voltage mismatch and intra-die leakage current in digital CMOS circuits," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 157-168, Jan. 2004.
- [30] Ming-Jer Chen, Jih-Shin Ho, and Tzuen-Hsi Huang, "Dependence of current match on back-gate bias in weakly inverted MOS transistor and its modeling," *IEEE J. Solid-State Circuits*, vol. 31, no. 2, pp. 259-262, Feb. 1996.
- [31] Jeroen A. Croon, Maarten Rosmeulen, Stefaan Decoutere, Willy Sansen, and Herman E. Maes, "An easy-to-use mismatch model for the MOS transistor," *IEEE J. Solid-State Circuits*, vol. 37, no. 8, pp. 1056-1064, Aug. 2002.
- [32] Shyh-Chyi Wong, Kuo-Hua Pan, and Dye-Jyun Ma, "A CMOS mismatch model and scaling effects," *IEEE Electron Device Lett.*, vol. 18, no. 6, pp. 261-263, June 1997.
- [33] R. Difrenza, P. Linares, E. Granger, H. Brut, G. Ghibaudo, "Effect of substrate voltage and oxide thickness on NMOSFET matching characteristics for a 0.18 μm CMOS technology," *Proc. IEEE ICMTS*, 2001, pp. 7-10.

Problems

6.1. Determine the expression that gives the mean square value of the offset voltage (mismatch of gate voltage for the same current) of a differential pair where the transistors operate in saturation. Plot the dependence of the deviation in the offset voltage on the inversion level i_f for $0.1 < i_f < 1000$. Assume the following parameters to calculate the offset deviation: $N_{oi} = 10^{12} \text{ cm}^{-2}$, $A_{ISQ} = 2\% \cdot \mu\text{m}$, $C'_{ox} = 10 \text{ fF}/\mu\text{m}^2$, $I_{SQ} = 100 \text{ nA}$, $n = 1.25$, $W = 10 \mu\text{m}$, $L = 1 \mu\text{m}$.

6.2. Assume that mismatch is due only to fluctuations in doping concentration. Let us examine the effects of technology scaling on matching, specifically in weak inversion, where the normalized current mismatch is given by (6.6.6). In the constant-field scaling (R. Dennard et al., "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE JSSC*, vol. 9, no. 5, pp. 256-268, Oct. 1974), dimensions are affected by k , the scaling factor, while doping is affected by $1/k$. Comment on the effect of technology scaling on the normalized error in weak inversion (Note: N_{oi} is proportional to $N_A^{1/2}$ where N_A is the doping level [7], [27]).

6.3. Assuming that mismatch is caused by random doping only, calculate the ratio of the normalized mismatch in saturation to the normalized mismatch in the linear region for $0.1 < i_f < 1000$.

This page intentionally left blank

Chapter 7

Noise in MOSFETs

This chapter deals with the fundamental noise in MOS transistors. The fundamental noise of a device is the result of the spontaneous fluctuations in current and voltage inside the device that are basically related to the discrete nature of electrical charge. Fundamental noise imposes limits on the performance of amplifiers and other electronic circuits. In contrast to pickup noise, that can be reduced or eliminated by appropriate shielding, fundamental noise cannot be eliminated because it is created in the device itself.

The study of noise in MOS transistors [1], [2] began shortly after the basic dc modeling had been carried out [3], [4] and was preceded by the study of noise in junction field-effect transistors [5], [6].

This chapter begins with a short summary of the various sources of noise. The impedance field method for calculating noise in devices is then introduced, and the fundamental thermal and flicker noise models are developed. Later, the basic requirements for a consistent noise model of the MOS transistor are given and design-oriented noise models are presented. Finally, small dimension effects on noise are considered.

7.1 Sources of noise

7.1.1 *Thermal noise*

Thermal noise is present in thermal equilibrium (in the absence of current) in any resistor. This is due to the fact that the electrons collide randomly with the thermally agitated atoms of the resistor. The effect is similar to the Brownian movement of particles suspended in a liquid

analyzed by Einstein in 1905 [7]. Brownian movement is basically explained by the equipartition theorem of statistical physics that states that for a system in equilibrium the average energy associated with each degree of freedom of the system is $kT/2$ where k is the Boltzmann constant and T is the temperature in Kelvin.

A suspended particle has three degrees of freedom and, consequently, three energy storage modes, each of which stores a mean energy of $kT/2$. For the electron gas in a (semi)conductor the situation is similar. In each storage mode in the circuit, such as a condenser or an inductor, the electron gas stores a mean energy of $kT/2$. This means that the total mean-square noise current in an inductor L is given by

$$\frac{1}{2} L \overline{i^2} = \frac{1}{2} kT. \quad (7.1.1)$$

Similarly, for a capacitor C , the mean-square noise voltage is

$$\frac{1}{2} C \overline{v^2} = \frac{1}{2} kT. \quad (7.1.2)$$

The thermal noise of a resistor has been determined applying the equipartition theorem to a lossless transmission line [8] or to a tuned circuit [9]. The thermal noise current of a resistor of value R ($=1/G$) has a mean-square value

$$\overline{i^2} = 4kT \frac{1}{R} \Delta f = 4kTG\Delta f, \quad (7.1.3)$$

where Δf is the bandwidth in Hertz over which the noise is measured. Since the noise current has a mean-square value that is proportional to the bandwidth, a noise current spectral density $\overline{i^2}/\Delta f$ that is constant as a function of the frequency can be defined. Noise with a constant spectral density is called white noise. Equation (7.1.3), which states that the thermal noise is directly proportional to the temperature T and to the conductance G was introduced and experimentally verified by Johnson [10], who measured the thermal noise in resistors made of different materials. Two equivalent circuits can represent the thermal noise of a resistor, as shown in Fig. 7.1 [11]. The Norton (shunt) equivalent circuit represents directly equation (7.1.3), while the Thevenin (series) equivalent circuit corresponds to expression (7.1.4)

$$\overline{v^2} = 4kTR\Delta f. \quad (7.1.4)$$

Equation (7.1.4) is very general because it is a consequence of thermodynamics, but it is strictly valid only in equilibrium, a condition that is only achieved with no current passing through the resistor. Because in many situations the electron drift velocities are much lower than the thermal velocity of the electron, (7.1.3) or (7.1.4) is also applied for biased resistors.

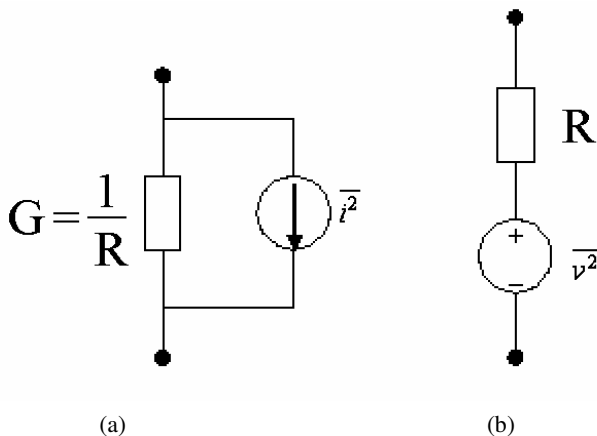


Fig. 7.1 (a) Norton and (b) Thevenin equivalent circuits of a real (noisy) resistor. The resistor R in the equivalent circuits is an ideal (noiseless) resistor.

7.1.2 Shot noise

Shot noise is associated with the random flow of carriers across a potential barrier. This noise is present in vacuum tubes, diodes and transistors. Shot noise was first studied by Schottky who explained the random fluctuations in the plate current as the result of the flow of discrete charges [12]. The fluctuation of the current I is given in terms of its mean-square variation about its average value I_{DC} as

$$\overline{i^2} = \overline{(I - I_{DC})^2}. \quad (7.1.5)$$

Shot noise is given by the Schottky formula as

$$\overline{i^2} = 2qI_{DC}\Delta f \quad (7.1.6)$$

where q is the electronic charge and Δf is the bandwidth in Hertz over which the noise is measured. As in the case of thermal noise, it is white noise.

Thermal noise and shot noise are related if the Einstein relation holds [13]. Let us consider the idealized diffusion current of electrons in one dimension (x -direction). Consider two adjacent boxes with cross-sectional areas equal to $\Delta y \Delta z$ and lengths equal to Δx . The electron concentrations in the two boxes are $n(x)$ and $n(x + \Delta x)$. Due to collisions with the semiconductor lattice, the electrons can jump randomly from one box to the adjacent box. The diffusion current I_F from left to right is proportional to the charge of carriers enclosed in the left box, *i.e.*,

$$I_F = -aqn(x) \Delta x \Delta y \Delta z \quad (7.1.7)$$

where a is a constant to be determined. The current I_R from right to left is

$$I_R = -aqn(x + \Delta x) \Delta x \Delta y \Delta z = -aq \left[n(x) + \frac{dn}{dx} \Delta x \right] \Delta x \Delta y \Delta z. \quad (7.1.8)$$

The net current I flowing from left to right is

$$I = I_F - I_R = aq\Delta x^2 \Delta y \Delta z \frac{dn}{dx}. \quad (7.1.9)$$

Since the current density $I/\Delta y \Delta z$ must be independent of the way in which the semiconductor is divided, $a\Delta x^2$ must be a constant; this constant is called the electron diffusion constant D_n . Therefore,

$$I = I_F - I_R = qD_n \frac{dn}{dx} \Delta y \Delta z. \quad (7.1.10)$$

Each of the currents is proportional to the local concentration of carriers $n(x)$, and the net diffusion current is proportional to dn/dx . For a constant concentration n , these two (forward and reverse) currents are equal, otherwise, considering that $\Delta n \ll n$, they are approximately equal, *i.e.*,

$$I_R \cong I_F = -qD_n n(x) \frac{\Delta y \Delta z}{\Delta x}. \quad (7.1.11)$$

The shot noise associated with each of these currents is given by (7.1.6)¹

$$\overline{i_{f(r)}^2} = 2qI_{F(R)}\Delta f = 2q[qD_n n(x) \frac{\Delta y \Delta z}{\Delta x}] \Delta f. \quad (7.1.12)$$

Considering that I_F and I_R are statistically independent, the mean square value of the noise is equal to the sum of the shot noise of each current. Thus, from (7.1.12) it follows that

$$\overline{i^2} = \overline{i_f^2} + \overline{i_r^2} = 4q^2 D_n n(x) \frac{\Delta y \Delta z}{\Delta x} \Delta f. \quad (7.1.13)$$

Now if the Einstein relation

$$D_n = \frac{kT}{q} \mu_n \quad (7.1.14)$$

holds, then (7.1.13) can be rewritten as

$$\overline{i^2} = 4kT[q\mu_n n(x)] \frac{\Delta y \Delta z}{\Delta x} \Delta f = 4kT \Delta G \Delta f, \quad (7.1.15)$$

which shows that the diffusion noise reduces to the thermal noise if the Einstein relation holds. ΔG is the conductance of the box ($\Delta x \Delta y \Delta z$).

7.1.3 Flicker noise

All active devices, as well as some passive devices such as carbon resistors, present what is referred to as excess noise, usually called flicker noise, in addition to thermal or shot noise. Flicker noise is a low-frequency noise which occurs when a direct current is flowing. The power spectral density (PSD) of flicker noise varies with frequency in the form

$$S(f) = \frac{K}{f^{EF}} \quad (7.1.16)$$

with K, EF being constants, $EF \approx 1$ (for a discussion on EF see [14]). Since usually $EF \approx 1$, as shown in Fig. 7.2, flicker noise is also called 1/f noise.

¹ In this section, $i_{f(r)}$ refers to noise current rather than inversion level.

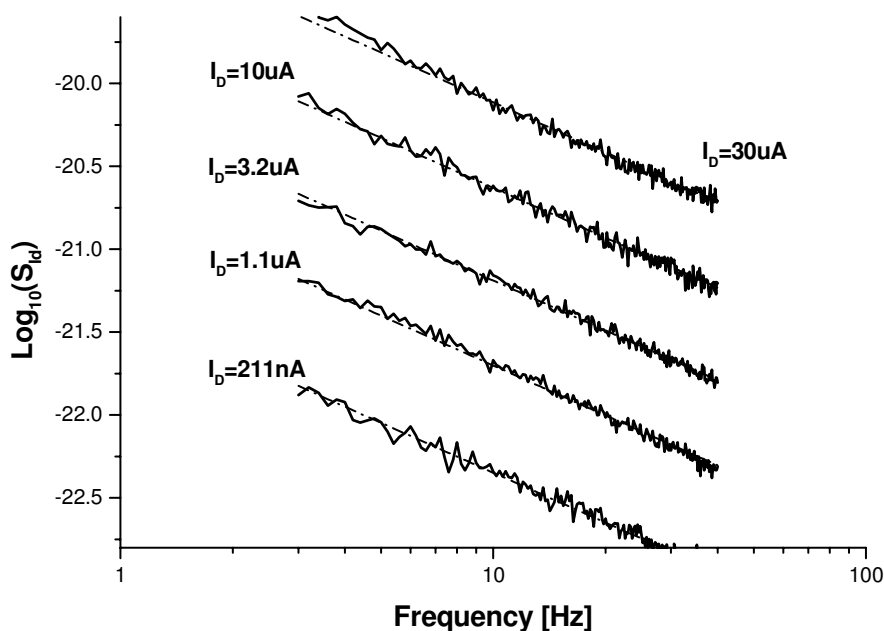


Fig. 7.2 Measured flicker noise spectra for a saturated NMOS transistor fabricated in a 2.4 μm process. $W/L=40\mu\text{m}/12\mu\text{m}$.

$1/f$ noise is ubiquitous; however, no universal mechanism, as in the case of thermal noise, is fully responsible for it. Consequently, $1/f$ noise cannot be predicted from the dc or other device characteristic as is the case with thermal or shot noise. Thus, empirical or semi-empirical parameters characterizing flicker noise must be determined from noise measurements. High $1/f$ noise, which has plagued MOS transistors from the beginning of the technology, has been studied since that time. However, there is still some controversy about its origin, even though nowadays the most accepted theory for $1/f$ noise in MOS transistors is its origin through interface traps, which will be the subject of section 7.2.2.

7.2 Noise modeling using the impedance field method

The impedance field method [15], [16] is the commonly used method for calculating the equivalent noise sources of a device from its structure

(technology and geometry) and the microscopic noise process involved (Fig. 7.3).

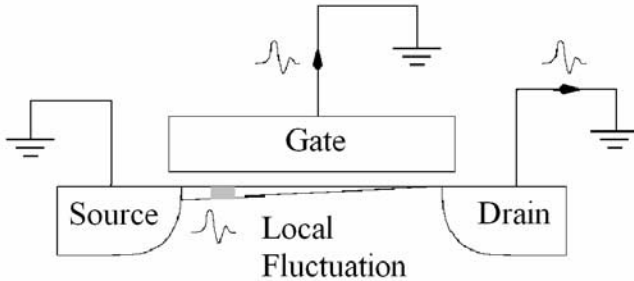


Fig. 7.3 A local fluctuation propagates to the gate and drain current of a MOSFET. (After [16].)

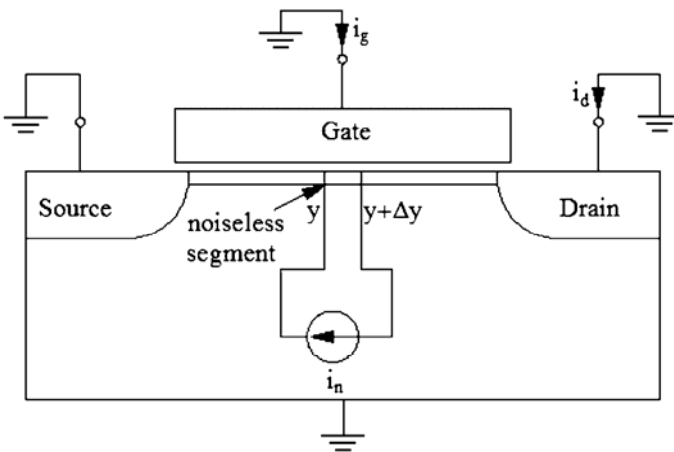


Fig. 7.4 The impedance field method applied to a MOSFET. (After [16].)

The calculation is carried out in three steps. Firstly, the local (microscopic) fluctuations are calculated in a generic elementary region of the device. A local current or voltage generator represents the local fluctuations (Fig. 7.4). Secondly, the transfer from the generic elementary region to the output voltage or current of the device is determined. The term impedance field is given owing to this procedure.

Since in our modeling approach we represent the local fluctuation by a current source and we are interested in the equivalent noise sources of the MOS, the impedance field in our case is given by a dimensionless current gain. Finally, the contributions of the elementary noise sources to the fluctuation of the output voltage or current must be combined. Since we applied the method to idealized one-dimensional structures, the transfer from the elementary region to the output is given in terms of the parameters of a generic lumped device. For these reasons, in the electrical engineering literature the impedance field method is sometimes simply called the circuit model approach [17].

The application of the impedance field method to the MOS transistor under the gradual channel approximation results in splitting the transistor into 3 series elements: the upper transistor, the lower transistor, and a small channel element of length Δy and area $\Delta A = W \cdot \Delta y$, as in Fig.7.5(a). Small signal analysis can be carried out, considering the general expression for the source (drain) transconductance $g_{ms(d)}$ given in (3.7.25). Thus

$$g_{msu} = -\mu \frac{W}{L-y} Q'_{IY} \quad (7.2.1)$$

$$g_{mdl} = -\mu \frac{W}{y} Q'_{IY} \quad (7.2.2)$$

where Q'_{IY} is the inversion charge density evaluated at a point Y in the channel and μ is the effective mobility. We define the resistance ΔR of a small element of the channel of length Δy with the aid of the general expression for the drain current, including drift and diffusion, as in equation (3.3.9), which is repeated here

$$I_D = -W \mu Q'_I \frac{dV_C}{dy}. \quad (7.2.3)$$

From (7.2.3), it follows that

$$\Delta R = \Delta V / I_D = -\Delta y / (\mu W Q'_I). \quad (7.2.4)$$

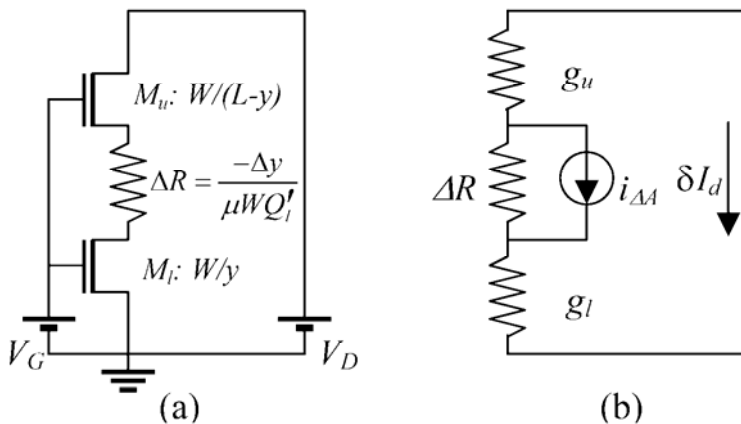


Fig.7.5 a) The transistor is separated into 3 series components. b) Small signal analysis to calculate the noise contribution to the drain current of the noisy element ΔA .

Representing the channel element by a resistance and considering the transconductance proportional to the inversion charge density in all operating regions are both consequences of the quasi-Fermi potential formulation for the drain current given by (7.2.3). Denoting the local noise current produced by the channel element as $i_{\Delta A}$ (Fig. 7.5 (b)), small-signal analysis allows one to calculate the resulting effect of $i_{\Delta A}$ on the drain current noise. As shown in Fig. 7.5(b), current division between the channel element and the equivalent small-signal resistance of the rest of the channel gives $\delta I_d = (\Delta y/L) i_{\Delta A}$. Thus, the mean-square of the total drain current fluctuation is²

$$\overline{\delta I_d^2} = \sum_{\text{channel-length}} \overline{(\delta I_d)^2} = \lim_{\Delta y \rightarrow 0} \sum \left(\frac{\Delta y}{L} i_{\Delta A} \right)^2 = \frac{1}{L^2} \int_0^L \Delta y \overline{i_{\Delta A}^2} dy \quad (7.2.5)$$

since local current fluctuations along the channel are uncorrelated. Note that since $i_{\Delta A}$ is related to the average of the local fluctuations of the carrier density N in the channel in the area ΔA , its PSD must be proportional to $1/\Delta A = 1/(W \cdot \Delta y)$. Equation (7.2.5) is general and may

² δ represents time fluctuations

be employed with any model for the noise $i_{\Delta A}$ of a single channel element such as the thermal noise model, Hooge's model [19], or the carrier number fluctuation model usually employed for the derivation of physics-based flicker noise models.

7.2.1 Thermal noise in MOSFETs

Calculating the mean-square thermal noise of a channel element of length Δy using (7.1.3), the Nyquist formula, gives

$$\frac{\overline{i^2}}{\Delta f} = 4kT \frac{W}{\Delta y} \mu(-Q'_I). \quad (7.2.6)$$

Substituting (7.2.6) into (7.2.5) yields

$$S_{th} = \frac{\overline{\delta I_D^2}}{\Delta f} = \frac{1}{L^2} \int_0^L 4kTW \mu(-Q'_I) dy = -4kT \mu \frac{Q_I}{L^2}, \quad (7.2.7)$$

where the rightmost term in (7.2.7) holds for the simple case of constant mobility. Equation (7.2.7), which states that the thermal noise is directly proportional to the total inversion charge Q_I , was first derived in [20].

From (7.2.7) and the expression for the total inversion charge in terms of the channel charge densities at the ends of the channel, as given in (5.2.16), it follows that

$$S_{th} = -4kT \mu \frac{W}{L} \frac{\frac{2}{3} (Q_{IS}'^2 + Q_{IS}' Q_{ID}' + Q_{ID}'^2) - nC'_{ox} \phi_t (Q_{IS}' + Q_{ID}')}{Q_{IS}' + Q_{ID}' - 2nC'_{ox} \phi_t}. \quad (7.2.8)$$

Expression (7.2.8) is valid in all the operating regions, from weak to strong inversion and from the linear to the saturation region, but is rather cumbersome. Useful design expressions in different operating regions in terms of the transistor transconductances are easily derived and will be presented in Section 7.4.1.

7.2.2 Flicker noise in MOSFETs

We will adopt the carrier number fluctuation theory [21], which is the most accepted approach to model the 1/f noise. Let us consider a small

channel element of area ΔA as in Fig. 7.6. Traps inside the oxide may be occupied by carriers from the channel. Electrons in the channel may tunnel to and back from these traps in a random process making the drain current fluctuate.

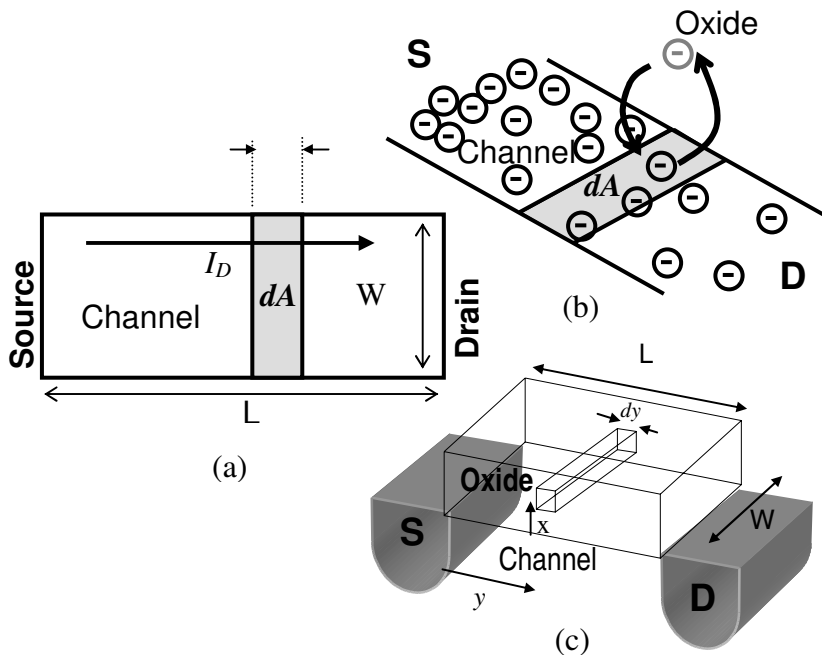


Fig. 7.6 (a) Top view of an NMOS transistor. (b) Electrons in the channel may tunnel to a trap. (c) Diagram of an MOS transistor and a differential volume inside the oxide $dV = W \cdot dx \cdot dy$.

Considering a small volume ΔV inside the oxide, as in Fig. 7.6(c): $\Delta V = W \cdot \Delta x \cdot \Delta y$, the following notation is introduced:

- $N'_{\Delta A}$ denotes the ratio of the carrier number in the channel element ΔA to the channel element area, which is also called areal carrier density.
- $N'_{t\Delta A}$ denotes the ratio of the number of occupied traps in the whole oxide volume above the channel element ΔA to the channel element area.
- $N_t(E)$ [$\text{eV}^{-1} \cdot \text{cm}^{-3}$] is the density of oxide traps per unit volume and unit energy.

To calculate the $1/f$ noise from (7.2.5), one must first determine $i_{\Delta A}$ through the following steps:

- 1) Find the relationship between the local noise current $i_{\Delta A}$ and the areal carrier density fluctuation $\delta N'_{\Delta A}$.
- 2) Determine the fluctuation $\delta N'_{\Delta A}$ in the areal carrier density in terms of the fluctuation in the areal density of occupied traps $\delta N'_{t\Delta A}$.
- 3) Finally, calculate $\delta N'_{t\Delta A}$.

The local current fluctuation that results from fluctuations in the areal carrier density is calculated from (7.2.3) as explained in [22], yielding

$$i_{\Delta A} = I_D \frac{\delta N'_{\Delta A}}{N'_{\Delta A}}. \quad (7.2.9)$$

Equation (6.4.2), equivalent to (7.2.9), has already been used in Section 6.4 to develop the transistor mismatch model. For the sake of simplicity we will consider only fluctuation in the number of carriers, but the analysis could be extended to include fluctuation in the mobility [23]. Following the same approach as that in Section 6.4, the ratio r of the fluctuation in areal carrier density to the fluctuation in the areal density of occupied traps [24] is given by the capacitive divider introduced in (6.4.4).

$$r = \frac{\delta N'_{\Delta A}}{\delta N'_{t\Delta A}} = -\frac{C'_i}{C'_i + C'_b + C'_{ox}} = \frac{Q'_I}{nC'_{ox}\phi_t - Q'_I}. \quad (7.2.10)$$

Finally, the power spectral density of the fluctuation $\delta N'_{t\Delta A}$ is calculated considering that the oxide traps have a uniform spatial distribution and that the probability of an electron penetrating into the oxide decreases exponentially with the distance from the interface. As shown in [14], under the above two hypotheses, the power spectral density of the fluctuation $\delta N'_{t\Delta A}$ is given by the very simple expression

$$\frac{(\delta N'_{t\Delta A})^2}{\Delta f} = \frac{1}{\Delta A} \frac{N_t kT}{\lambda} \frac{1}{f} = \frac{N_{ot}}{\Delta A} \frac{1}{f}. \quad (7.2.11)$$

Parameter N_{ot} is the equivalent density of oxide traps defined by

$$N_{ot} [\text{cm}^{-2}] = \frac{kTN_t(E)}{\lambda}, \quad (7.2.12)$$

where λ (cm^{-1}) is the attenuation coefficient of the electron wave function in the oxide.

From (7.2.10) and (7.2.11), it follows that the PSD of the fluctuations of the areal carrier density $\delta N'_{\Delta A}$ is

$$\frac{(\delta N'_{\Delta A})^2}{\Delta f} = \frac{N_{ot} \cdot (Q'_I)^2}{\Delta A \cdot (Q'_I - nC'_{ox}\phi_t)^2} \cdot \frac{1}{f}. \quad (7.2.13)$$

We obtain an expression for the PSD of the drain current noise using (7.2.9) and (7.2.13) to calculate $i_{\Delta A}$, which inserted into (7.2.5), yields

$$\frac{\overline{\delta I_D^2}}{\Delta f} = \frac{q^2 N_{ot} I_D^2}{WL^2} \frac{1}{f} \int_0^L \frac{1}{(nC'_{ox}\phi_t - Q'_I)^2} dy. \quad (7.2.14)$$

With the aid of (5.2.3) the integration over the channel length in (7.2.14) is changed into the integration over the channel charge density. Thus,

$$\frac{\overline{\delta I_D^2}}{\Delta f} = \frac{q^2 N_{ot} \mu I_D}{nC'_{ox} L^2} \cdot \frac{1}{f} \int_{Q'_{IS}}^{Q'_{ID}} \frac{1}{nC'_{ox}\phi_t - Q'_I} dQ'_I. \quad (7.2.15)$$

The integration of (7.2.15) results in

$$\frac{\overline{\delta I_D^2}}{I_D^2 \Delta f} = \frac{q^2 N_{ot} \mu}{L^2 nC'_{ox} I_D} \ln \left[\frac{nC'_{ox}\phi_t - Q'_{IS}}{nC'_{ox}\phi_t - Q'_{ID}} \right] \frac{1}{f}. \quad (7.2.16)$$

An expression similar to (7.2.16) is used in BSIM [23] to model strong inversion noise due to charge trapping, but it must be emphasized that (7.2.16) is valid for any inversion level, including moderate inversion.

Approximate expressions can be found for the different operating regions. In weak inversion, $-Q'_{ID}, -Q'_{IS} \ll nC'_{ox}\phi_t$. Making a first order series expansion, it is possible to rewrite (7.2.16) in a more concise manner as

$$\frac{S_{I_d}}{I_D^2} = \frac{N_{ot}}{WLN^{*2}} \cdot \frac{1}{f} \quad (7.2.17)$$

where $N^* = nC'_{ox}\phi_t/q$ [23]. Equation (7.2.17) is the same one used in BSIM3 and 4 to model flicker noise in weak inversion ($N_{ot} = AkT/\lambda$ in BSIM3, 4) and it is equivalent to the formula proposed in [24].

In the linear region in strong inversion, $Q'_{ID} \cong Q'_{IS} \cong -C'_{ox}(V_G - V_T)$, and the first order expansion of (7.2.16) leads to

$$\frac{S_{I_d}}{I_D^2} = \frac{q^2 N_{ot}}{WLC'^2_{ox}(V_G - V_T)^2} \cdot \frac{1}{f}. \quad (7.2.18)$$

In this case, the result is also the same as that obtained in [24]. Summarizing the bias dependence of the PSD of the drain current noise, in subthreshold the PSD increases with I_D^2 according to (7.2.17) while in strong inversion saturation the PSD increases linearly with I_D , as predicted by (7.2.16) if one neglects the variation in the logarithmic term.

7.2.2.1 Measurement results

Noise measurements were performed in MOS transistors covering all regions of operation to experimentally verify the model in (7.2.16). In the experiments, the PSD closely follows a $1/f$ dependence. This is consistent with the assumption $EF = 1$ associated with a uniform spatial distribution of the traps inside the oxide [14]. Figures 7.7 and 7.8 show the simulation and measurement of flicker noise for an n-channel MOSFET of a $0.8\mu\text{m}$ CMOS process. Fig. 7.7 covers all the operating regions for the saturated transistor, from weak to strong inversion; note here the plateau of the normalized PSD S_{I_d}/I_D^2 in weak inversion. Fig. 7.8 was obtained for the same transistor operating in strong inversion and in the linear region. Both transistors in Fig. 7.8 and Fig. 7.9 were fabricated in the same technology but in different runs, which may explain the slightly different values of N_{ot} .

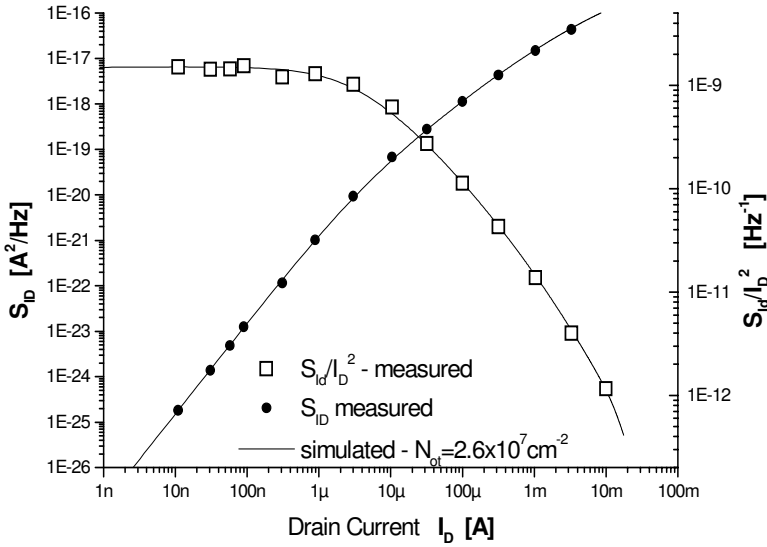


Fig. 7.7 Flicker noise PSD and normalized PSD at $f=1$ Hz, for a saturated NMOS transistor ($W/L=200\mu\text{m}/5\mu\text{m}$).

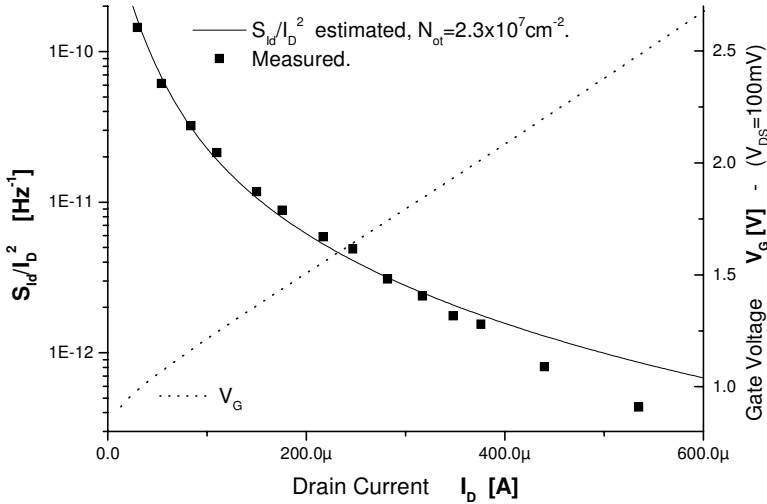


Fig. 7.8 Normalized PSD at $f=1$ Hz, for an NMOS transistor ($W/L=200\mu\text{m}/5\mu\text{m}$) in the linear region. (The drain-to-source voltage is 100 mV .)

In Fig. 7.9 the predicted and measured flicker noise for a $W/L=20\mu\text{m}/10\mu\text{m}$ NMOS transistor covering different V_{DS} values from the linear region up to saturation with a gate voltage $V_G=3.3\text{V}$ are shown.

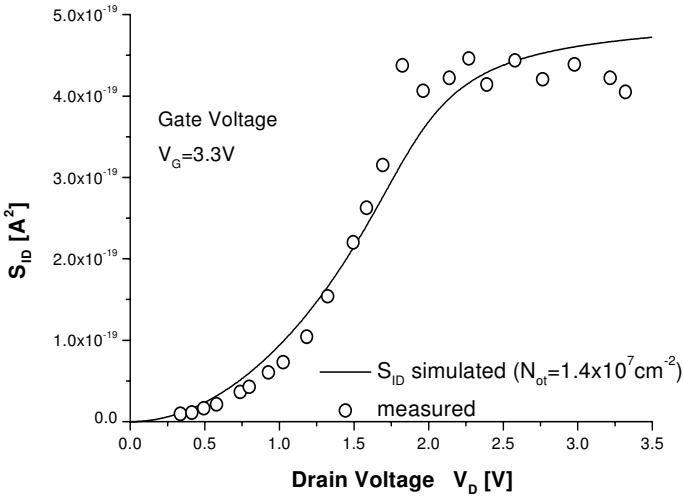


Fig. 7.9. PSD of flicker noise at $f=1$ Hz for a $W/L=20\mu\text{m}/10\mu\text{m}$ NMOS transistor, from the linear region up to saturation.

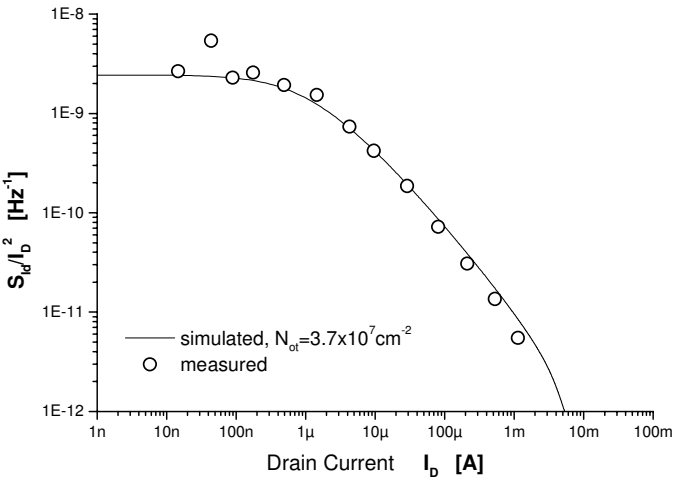


Fig. 7.10 Normalized PSD of the drain current noise at $f=1$ Hz for a saturated PMOS transistor ($W/L=200\mu\text{m}/5\mu\text{m}$).

In Fig. 7.10 the flicker noise is shown for a saturated p-channel transistor of the same technology and with the same aspect ratio as the one in Fig. 7.9. As shown in Fig. 7.10, it is also possible to observe the plateau in weak inversion.

Finally, in Fig. 7.11 measurements and simulations of flicker noise for a 2.4 μm technology are presented. In this case the measurements were performed for two NMOS transistors of different sizes. Once again the measurements show a good agreement with the model and the value of the extracted N_{ot} is similar to the values determined for the 0.8 μm process.

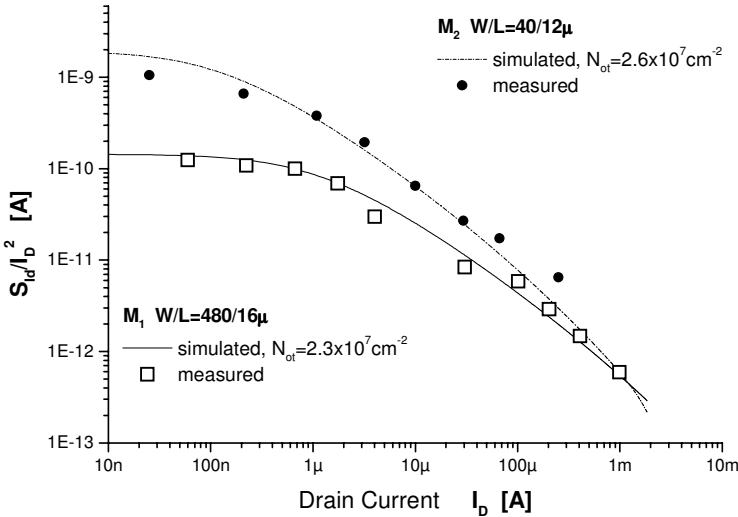


Fig. 7.11 Normalized PSD of the drain current noise at $f=1$ Hz for two saturated NMOS transistors in a 2.4 μm CMOS process.

7.3 Consistency of noise models

A noise model is consistent regarding series or parallel associations if the composition of the noise contributions from the individual series (or parallel) elements is the same as the noise from the series (or parallel) equivalent. Clearly, the thermal noise model for a resistor, as in equation (7.1.3), repeated here

$$\frac{\overline{i^2}}{\Delta f} = \frac{4kT}{R} \quad (7.3.1)$$

is consistent [11]. For two series elements R_1, R_2 (Fig. 7.12) the total noise current introduced into the circuit: $\overline{i^2} / \Delta f = 4kT / (R_1 + R_2)$ can be obtained by composing the individual noise sources or using (7.3.1) to calculate the noise of the equivalent resistor $R_{eq} = R_1 + R_2$.

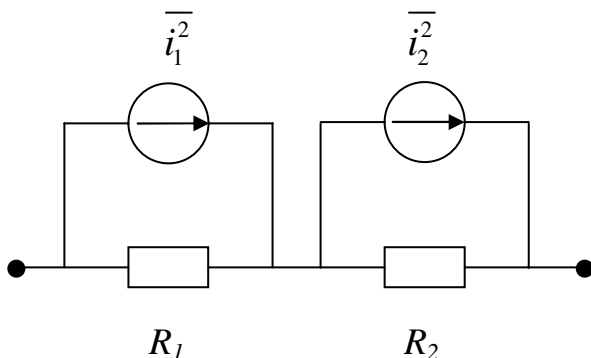


Fig. 7.12 Circuit for the calculation of total noise produced by two resistors in series.

The analysis can be extended to MOS transistors, because for these devices, series and parallel equivalents are clearly defined [25], [26]. Consider, for example, the virtual cut of a transistor that slices it into two series elements as in Fig. 7.13. Suppose that the upper and lower transistors M_u and M_l introduce noise currents with PSD values equal to $S_{i_{du}}$ and $S_{i_{dl}}$, respectively. Small-signal analysis (see Fig. 7.13(c)) allows the calculation of the PSD of the noise current $S_{i_{d_s}}$ of the series-composed transistor. Considering $S_{i_{dl}}$ and $S_{i_{du}}$ to be the PSD of uncorrelated noise current sources, it follows that

$$S_{i_{d_s}}(f) = \left[\frac{a}{1+a} \right]^2 S_{i_{dl}}(f) + \left[\frac{1}{1+a} \right]^2 S_{i_{du}}(f), \quad (7.3.2)$$

where $a = g_{msu}/g_{mdl}$, and g_{msu} and g_{mdl} are the source and drain transconductances of transistors M_u and M_l , respectively. For the partition of the channel as in Fig. 7.13 we have

$$g_{msu} = -\mu \frac{W}{L-d} Q'_{IY} \quad g_{mdl} = -\mu \frac{W}{d} Q'_{IY} \quad (7.3.3)$$

where Q'_{IY} is the inversion charge density evaluated at point Y in the channel (Fig. 7.13(b)) and μ is the effective mobility. Consequently, $a = d/(L-d)$ depends only on the position in the channel and (7.3.2) can be rewritten as

$$S_{i_{d_s}}(f) = \left[\frac{d}{L} \right]^2 S_{i_{dl}}(f) + \left[\frac{L-d}{L} \right]^2 S_{i_{du}}(f). \quad (7.3.4)$$

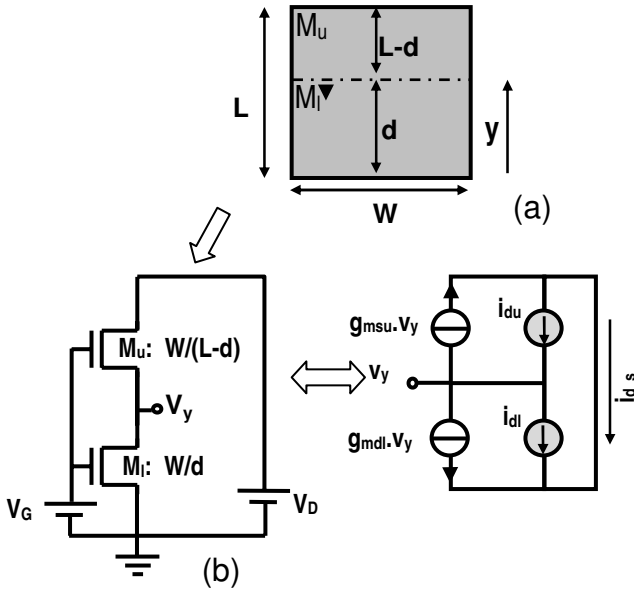


Fig. 7.13 Circuit for the calculation of the total noise produced by two transistors in series.

As an example, let us now consider the application of (7.3.4) to thermal noise. Calculating the PSD of the upper and lower transistors

using the MOSFET thermal noise formula (7.2.7) and substituting the result into (7.3.4), yields:

$$S_{th} = -4kT\mu \left[\frac{Q_{ll}}{d^2} \left(\frac{d}{L} \right)^2 + \frac{Q_{lu}}{(L-d)^2} \left(\frac{L-d}{L} \right)^2 \right] = \frac{-4kT\mu Q_I}{L^2} \quad (7.3.5)$$

where Q_{ll} , Q_{lu} , Q_I are the total inversion charge in the channel of the lower, upper, and equivalent transistor, respectively. As expected, the classical thermal noise model of the MOSFET is consistent with the series association of transistors.

The flicker noise model of (7.2.16) was deduced from a charge-based model integrated along the transistor channel, thus resulting in an inherently consistent model for the series and parallel associations of transistors. However, some popular 1/f noise models are inconsistent, as pointed out in [14], [27].

7.4 Design-oriented noise models

7.4.1 Relationship between thermal noise and transconductance

In the linear region from weak to strong inversion $Q'_{IS} \cong Q'_{ID}$ and

$$S_{th} \cong -4kT\mu \frac{W}{L} Q'_{IS} = 4kT g_{ms} \cong 4kT g_{md}. \quad (7.4.1)$$

As expected, the channel behaves as a resistance of value $1/g_{ms}=1/g_{md}$. In weak inversion, $|Q'_{IS(D)}| \ll nC'_{ox}\phi_t$, and (7.2.8) becomes

$$S_{th} \cong -4kT\mu \frac{W}{L} \frac{(Q'_{IS} + Q'_{ID})}{2} = 4kT \frac{g_{ms} + g_{md}}{2}. \quad (7.4.2)$$

For a saturated transistor ($g_{ms} \gg g_{md}$) in weak inversion

$$S_{th} = 2nkTg_m. \quad (7.4.3)$$

In saturation $|Q'_{IS}| \gg |Q'_{ID}|$ and, in strong inversion, $|Q'_{IS}| \gg nC'_{ox}\phi_t$. Thus, it is possible to rewrite (7.2.8) as

$$S_{th} = \frac{8}{3} nkTg_m. \quad (7.4.4)$$

In Fig. 7.14 the calculated and measured values of the normalized PSD of white noise are shown. These measurements were obtained at a frequency of several kHz to minimize the effect of flicker noise.

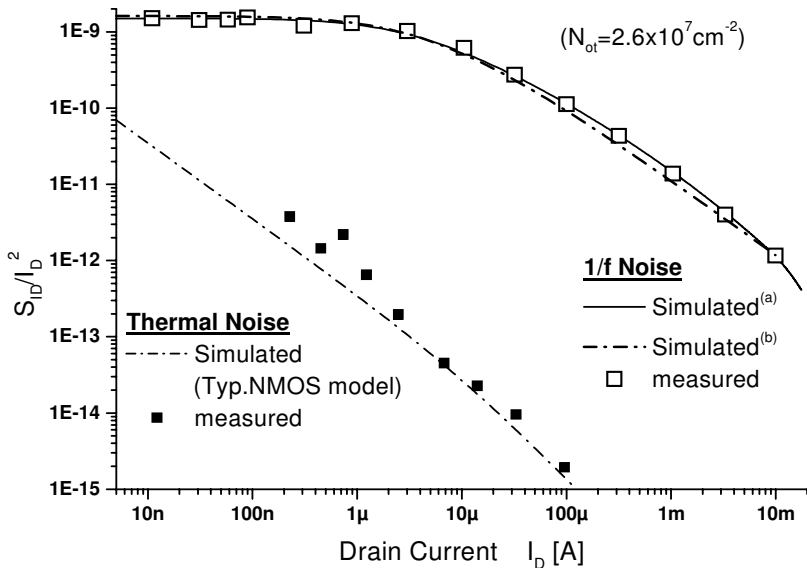


Fig. 7.14 Normalized thermal noise PSD and normalized 1/f noise PSD for a saturated NMOS ($W/L=200\mu/5\mu$) transistor. The simulated results for 1/f noise were obtained through (a) Equation (7.2.16), (b) Expression (7.4.12) times g_m^2 .

7.4.2 Flicker noise in terms of inversion levels

A useful alternative expression for (7.2.16) is obtained if the charge densities at source (drain) are expressed in terms of the normalized forward and reverse currents i_f and i_r . Using the relationship between normalized charges and currents of (3.7.3), expression (7.2.16) can be rewritten as

$$\frac{S_{I_D}}{I_D^2} = \frac{N_{ot}}{WLN^{*2}} \frac{1}{f} \frac{1}{(i_f - i_r)} \ln \left[\frac{1+i_f}{1+i_r} \right]. \quad (7.4.5)$$

From weak to strong inversion in the linear region, $i_f \approx i_r$ and (7.4.5) reduces to

$$\frac{S_{I_d}}{I_D^2} = \frac{N_{ot}}{WLN^{*2}} \frac{1}{f} \frac{1}{1+i_f}. \quad (7.4.6)$$

In weak inversion, $i_f \ll 1$ and $i_r \ll 1$. The first-order series expansion of (7.4.5) leads to

$$\frac{S_{I_d}}{I_D^2} = \frac{N_{ot}}{WLN^{*2}} \frac{1}{f}. \quad (7.4.7)$$

Sometimes, designers prefer to represent the channel noise as an equivalent noise voltage source in series with the gate rather than as a noise current source connected between drain and source. In this case, the PSD of the noise voltage source in series with the gate is

$$S_{V_g} = \frac{S_{I_d}}{g_m^2}. \quad (7.4.8)$$

The ratio of the drain current to the gate transconductance in terms of the inversion level for a transistor operating in saturation, given by (3.7.31), is repeated below

$$\frac{I_D}{n\phi_t g_m} = \frac{1 + \sqrt{1+i_f}}{2}. \quad (7.4.9)$$

Thus, from (7.4.5) and (7.4.9) it follows that

$$S_{V_g} = \frac{q^2 N_{ot}}{WLC_{ox}^2} \frac{1}{f} \psi(i_f) \quad (7.4.10)$$

where

$$\psi(i_f) = \left(\frac{1 + \sqrt{1+i_f}}{2} \right)^2 \frac{\ln(1+i_f)}{i_f}. \quad (7.4.11)$$

Because $\psi(i_f)$ shows very small variation with i_f as seen in Fig. 7.15, the following empirical model results if we consider ψ equal to 1:

$$S_{V_g} \cong \frac{q^2 N_{ot}}{WLC'_{ox}} \frac{1}{f}. \quad (7.4.12)$$

Expression (7.4.12) is sometimes presented as

$$S_{V_g} \cong \frac{K_F}{WLC'_{ox}} \frac{1}{f}, \quad (7.4.13)$$

where the flicker noise constant $K_F = q^2 N_{ot} / C'_{ox}$.

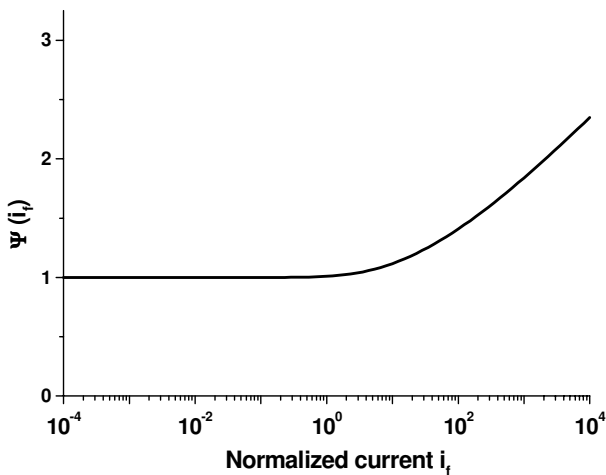


Fig. 7.15 Function $\psi(i_f)$ versus normalized forward current.

Due to its simplicity, the empirical model (7.4.12) is very convenient for hand calculations. Moreover, in current designs, the inversion level is seldom higher than 10^2 , thus $\psi \approx 1$ is a fair approximation. Even though the empirical model of (7.4.12) gives a good estimate of the flicker noise of a transistor in saturation, it is not consistent with expression (7.3.4). In effect, the empirical model does not consider the distributed nature of the MOSFET, because it represents noise as a gate voltage source independent of the bias condition ($\psi(i_f) = 1$). It should be noted that for

high current applications such as those in RF circuits, equation (7.4.12) can give large errors. For a gate overdrive $V_{GS}-V_T$ of 1.5 V, $i_f \approx 2000$ and $\psi \approx 2$.

7.4.3 The corner frequency

The corner frequency f_c , defined as the frequency at which the flicker noise and thermal noise PSD have the same value, can be calculated directly in terms of Q'_{IS}, Q'_{ID} from equations (7.2.8) and (7.2.16). However, simpler results were obtained determining f_c in weak inversion with equations (7.4.3) and (7.2.17) and in strong inversion using (7.4.4) and (7.4.5)

$$f_c = \frac{\alpha g_m}{WLC'_{ox}} \frac{N_{ot}}{N^*} \cong \frac{\pi}{2} \frac{N_{ot}}{N^*} f_T \quad (7.4.14)$$

with $\alpha = 1/2$ in weak inversion and $\alpha \cong 9/16$ in strong inversion. Note that the corner frequency in (7.4.14) is proportional to the transition frequency f_T of the transistor, which is a useful approximation for designers.

The total noise in a frequency band (f_2-f_1) resulting from the contributions of thermal and flicker noise can be calculated as an equivalent voltage referred to the gate. For a saturated transistor operating in weak inversion, the integration of both (7.2.17) and (7.4.3), yields

$$\overline{v_g^2} = \frac{2nkT}{g_m} \left((f_2 - f_1) + f_c \ln \left(\frac{f_2}{f_1} \right) \right). \quad (7.4.15)$$

For strong inversion, an analogous formula exists with slightly different coefficients. Fig. 7.16 shows the simulated and measured corner frequency of a saturated NMOS transistor for various bias currents. The solid line represents f_c calculated using (7.4.14), together with the measured value for g_m . The dashed line represents f_c calculated using the expression of g_m in terms of the inversion level. For this transistor, the dimensionless factor $N_{ot}/N^* \cong 0.5 \times 10^{-3}$.

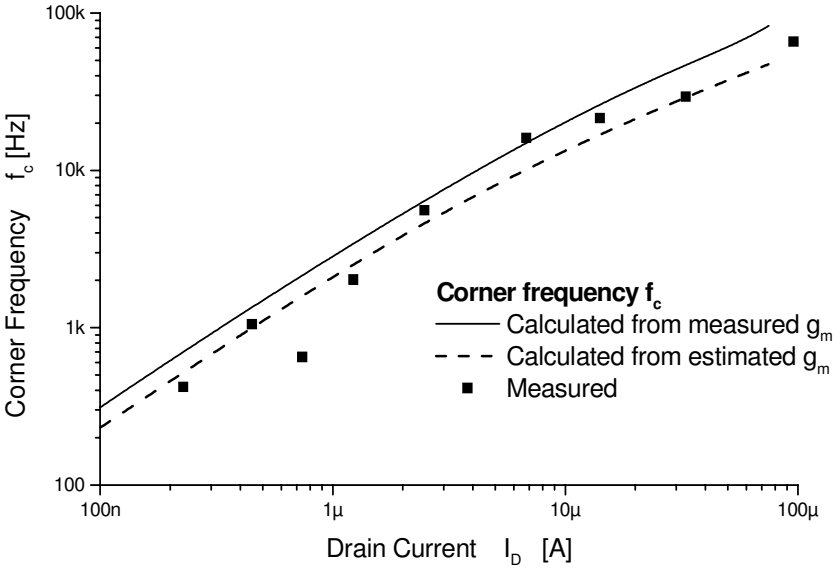


Fig. 7.16 Calculated and measured values of the corner frequency f_c , for an NMOS transistor with $W/L=200\mu\text{m}/5\mu\text{m}$.

7.5 Small-dimension effects on MOSFET noise

This section presents the derivation of a closed expression for the channel thermal noise including the velocity saturation effect in all the operating regions of the MOSFET. The procedure used for the derivation follows that used in [28], but rather than constraining the result to strong inversion, the solution obtained is quite general and is expressed in terms of the inversion charge densities at the source and drain terminals.

7.5.1 Review of the effect of velocity saturation on the drain current

This subsection reviews some dc equations required for the deduction of the thermal noise in short-channel MOSFETs.

The relationship between inversion charge density Q'_i and channel potential V_C is given by the unified charge control model (UCCM), which in normalized form is given by (3.7.6) and rewritten below

$$\frac{V_P - V_C}{\phi_t} = q'_I - 1 + \ln(q'_I), \quad (7.5.1)$$

where $q'_I = Q'_I / Q'_{IP}$. From (7.5.1) we find that

$$-\frac{dV_C}{dy} = \phi_t \left[1 + \frac{1}{q'_I} \right] \frac{dq'_I}{dy}. \quad (7.5.2)$$

To account for the mobility dependence on the longitudinal field, we write the expression for the mobility as

$$\mu = \frac{\mu_s}{1 - \frac{F}{F_c}} = \frac{\mu_s}{1 - \frac{\mu_s F}{v_{sat}}}, \quad (7.5.3)$$

which results in the following expression for the drain current (see (4.2.5)), here written in normalized form

$$i_D = \frac{I_D}{I_S} = \frac{-2}{1 - \zeta \frac{dq'_I}{d\xi}} [q'_I + 1] \frac{dq'_I}{d\xi}, \quad (7.5.4)$$

where

$$I_S = \frac{\mu_s C'_{ox} n \phi_t^2}{2} \frac{W}{L} \quad (7.5.5)$$

is the specific current for low-field mobility, $\xi = y/L$ is the normalized distance to the source, and

$$\zeta = \frac{\phi_t}{L F_c} = \frac{\mu_s \phi_t}{L v_{sat}} \quad (7.5.6)$$

is the short-channel factor. The integration of (7.5.4) along the transistor channel, from source to drain results in (4.2.9), repeated below

$$i_D = \frac{(q'_{IS} + q'_{ID} + 2)}{1 + \zeta (q'_{IS} - q'_{ID})} (q'_{IS} - q'_{ID}) \quad (7.5.7)$$

for the drain current normalized to the specific current I_S . The normalized inversion charge densities q'_{IS} and q'_{ID} in (7.5.7) are calculated at source and drain, respectively.

7.5.2 The impedance field for short channel MOSFETs

In order to calculate the channel noise, we now proceed as in [28], with local noise represented by two current sources, according to the representation in Fig. 7.17. The effect of each current source on the drain current is calculated using the impedance field method [15]. The channel conductances G_1 and G_2 are determined from

$$G_{1(2)} = \frac{dI_{1(2)}}{dV_C} = I_s \frac{di_{1(2)}}{dV_C}. \quad (7.5.8)$$

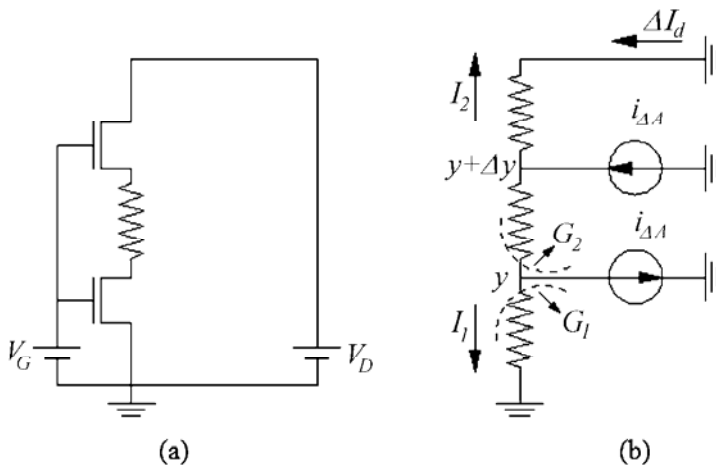


Fig. 7.17 Transistor partition for calculating the contribution of elementary noise sources [28].

Since the dependence of the current on the inversion charge is known and the relationship between charge and channel potential is given by UCCM, we can solve (7.5.8) by using expressions (7.5.4) and (7.5.1). Thus,

$$G_{1(2)} = G_0 \frac{di_{1(2)}}{dq'_I} \frac{dq'_I}{dV_C/\phi_I} = -G_0 \frac{di_{1(2)}}{dq'_I} \frac{q'_I}{q'_I + 1}, \quad (7.5.9)$$

where $G_0 = I_s/\phi_I$ is the normalization conductance.

The derivatives of the currents wrt the inversion charge are calculated after we have computed the expressions for $i_{1(2)}$ by integrating expression

(7.5.4) between the source (infinitesimal channel element) and the infinitesimal channel element (drain), which results in

$$i_1 = \frac{(q'_{IS} + 1)^2 - (q'_I + 1)^2}{\xi + \zeta (q'_{IS} - q'_I)} \quad (7.5.10)$$

for i_1 and

$$i_2 = -\frac{(q'_I + 1)^2 - (q'_{ID} + 1)^2}{(1 - \xi) + \zeta (q'_I - q'_{ID})} \quad (7.5.11)$$

for i_2 . The derivatives of the currents wrt the local charge become

$$\frac{di_1}{dq'_I} = \frac{-2(q'_I + 1) + \zeta i_1}{\xi + \zeta (q'_{IS} - q'_I)}, \quad (7.5.12)$$

and

$$\frac{di_2}{dq'_I} = \frac{-2(q'_I + 1) + \zeta i_2}{(1 - \xi) + \zeta (q'_I - q'_{ID})}. \quad (7.5.13)$$

Note that the dc current $i_1 = -i_2$. Now, to obtain the effect of the elementary channel noise $i_{\Delta A}$ on the drain current ΔI_d , one superimposes the effects of the two elementary current sources to give

$$\Delta I_d = \left(\frac{G_2}{G_1 + G_2} \Big|_{\xi + \Delta \xi} - \frac{G_2}{G_1 + G_2} \Big|_{\xi} \right) i_{\Delta A}. \quad (7.5.14)$$

The substitution of (7.5.12) and (7.5.13) into (7.5.9) leads to the following result for the current division

$$\frac{G_2}{G_1 + G_2} = \frac{\xi + \zeta (q'_{IS} - q'_I)}{1 + \zeta (q'_{IS} - q'_{ID})}. \quad (7.5.15)$$

The result in (7.5.15) is a generalization of the current division principle, derived under the assumption that the mobility follows the dependence on the longitudinal field as given by (7.5.3). For long channel devices, the current division simply states that, for a given current source injected into the channel, the fraction of the current that flows through the drain is proportional to the distance from the injection node to the source end of the channel. If, however, the saturation velocity

phenomenon is significant, the current division principle should be modified according to expression (7.5.15), derived under the assumption of a mobility dependence on longitudinal field according to expression (7.5.3). At this point, let us apply (7.5.15) to a transistor operating in strong inversion, where $q'_I \gg 1$. Therefore, $q'_I \cong (V_p - V_C)/\phi_t$, which leads to the following de-normalized relationship for the current divider

$$\frac{G_2}{G_1 + G_2} \cong \frac{y + \frac{V_{CS}}{F_c}}{L + \frac{V_{DS}}{F_c}}, \quad (7.5.16)$$

where V_{CS} is the channel-to-source voltage. The result in (7.5.16) is equal to that in [28].

Using expression (7.5.15) in (7.5.14) results in

$$\Delta I_d = \frac{\Delta \xi - \zeta \Delta q'_I}{1 + \zeta (q'_{IS} - q'_{ID})} i_{\Delta A} = \frac{\Delta \xi \left(1 - \zeta \frac{dq'_I}{d\xi} \right)}{1 + \zeta (q'_{IS} - q'_{ID})} i_{\Delta A}. \quad (7.5.17)$$

7.5.3 Drain current noise of short channel MOSFETs

In order to calculate the total drain current noise, recall that the diffusion noise, *i.e.*, the noise caused by the collision of carriers with the lattice, is given [13] by

$$\overline{i_{\Delta A}^2} = -4qD_n Q'_I(y) \frac{W}{\Delta y} \Delta f \quad (7.5.18)$$

for an infinitesimally small thermal noise source located between y and $y + \Delta y$. In (7.5.18), q is the electronic charge and D_n is the carrier diffusion coefficient. In covalent semiconductors such as silicon, the diffusion constant decreases very slowly with increasing field strength [29], [30]. For the sake of compactness in the derivation of the noise model, we will assume simply that the diffusion coefficient is independent of the electrical field. The use of a constant diffusion

coefficient in (7.5.18) implicitly accounts for the hot carrier case in which the carriers are heated by the electrical field [28].

To normalize equation (7.5.18), the constant diffusion coefficient can be expressed in terms of the low-field mobility, as given by the Einstein relationship $D_n = D_0 = \mu_0 \phi_t$. Rewriting (7.5.18) in normalized form, we then obtain

$$\frac{\overline{i_{\Delta A}^2}}{4kT\Delta f} = \frac{2G_0}{\Delta \xi} q'_I. \quad (7.5.19)$$

We can proceed now to determine the total drain current noise by adding the contribution of all channel elements, *i.e.*

$$I_d = \sum_{channel} \Delta I_d. \quad (7.5.20)$$

In order to evaluate the mean-square value of the drain current, we calculate the following summation along the channel

$$\overline{I_d^2} = \overline{\left(\sum_{channel} \Delta I_d \right)^2}. \quad (7.5.21)$$

Using (7.5.17) in (7.5.21) we find that

$$\overline{I_d^2} = \overline{\left(\sum_{channel} \frac{\Delta \xi \left(1 - \zeta \frac{dq'_I}{d\xi} \right)}{1 + \zeta (q'_{IS} - q'_{ID})} i_{\Delta A} \right)^2}. \quad (7.5.22)$$

If the elementary thermal noise sources are assumed to be uncorrelated, then (7.5.22) gives

$$\overline{I_d^2} = \sum_{channel} \left(\frac{\Delta \xi \left(1 - \zeta \frac{dq'_I}{d\xi} \right)}{1 + \zeta (q'_{IS} - q'_{ID})} \right)^2 \overline{i_{\Delta A}^2}. \quad (7.5.23)$$

Substituting expression (7.5.19) into (7.5.23) leads to

$$\frac{\overline{I_d^2}}{4kTG_0\Delta f} = \sum_{channel} 2 \frac{\Delta\xi \left(1 - \zeta \frac{dq'_I}{d\xi}\right)^2}{\left[1 + \zeta (q'_{IS} - q'_{ID})\right]^2} q'_I. \quad (7.5.24)$$

Making $\Delta\xi \rightarrow 0$, the summation along the channel becomes

$$\frac{\overline{I_d^2}}{4kTG_0\Delta f} = 2 \frac{\int_{source}^{drain} q'_I \left(1 - \zeta \frac{dq'_I}{d\xi}\right)^2 d\xi}{\left[1 + \zeta (q'_{IS} - q'_{ID})\right]^2}, \quad (7.5.25)$$

and using (7.5.4) we find that

$$\frac{\overline{I_d^2}}{4kTG_0\Delta f} = -\frac{4}{i_D} \frac{\int_{source}^{drain} q'_I (q'_I + 1) \left(1 - \zeta \frac{dq'_I}{d\xi}\right) dq'_I}{\left[1 + \zeta (q'_{IS} - q'_{ID})\right]^2}. \quad (7.5.26)$$

The integral in (7.5.26) can be separated into two terms as in [28], the first corresponding to thermal noise under equilibrium (the Einstein relationship is assumed to hold for strong electric fields) while the second corresponds to noise added by the heating effect, *i.e.*,

$$\overline{I_d^2} = \overline{I_{d,equ}^2} + \overline{I_{d,heat}^2}, \quad (7.5.27)$$

where

$$\frac{\overline{I_{d,equ}^2}}{4kTG_0\Delta f} = -\frac{4}{i_D} \frac{\int_{source}^{drain} q'_I (q'_I + 1) dq'_I}{\left[1 + \zeta (q'_{IS} - q'_{ID})\right]^2} \quad (7.5.28)$$

and

$$\frac{\overline{I_{d,heat}^2}}{4kTG_0\Delta f} = \frac{4\epsilon}{i_D} \frac{\int_{source}^{drain} q'_I (q'_I + 1) \frac{dq'_I}{d\xi} dq'_I}{\left[1 + \zeta (q'_{IS} - q'_{ID})\right]^2}. \quad (7.5.29)$$

The integration of (7.5.29) along the channel gives for the mean square value of the drain current

$$\frac{\overline{I_{d, equ}^2}}{4kTG_0\Delta f} = 4 \frac{\frac{(q'_{IS} + 1)^3 - (q'_{ID} + 1)^3}{3} - \frac{(q'_{IS} + 1)^2 - (q'_{ID} + 1)^2}{2}}{i_D [1 + \zeta (q'_{IS} - q'_{ID})]^2}. \quad (7.5.30)$$

Expression (7.5.7) of the drain current allows us to rewrite (7.5.30) as

$$\frac{\overline{I_{d, equ}^2}}{4kTG_0\Delta f} = 2 \frac{\left[\frac{2 (q'_{IS} + 1)(1 + \eta + \eta^2)}{3(1 + \eta)} - 1 \right]}{1 + \zeta (q'_{IS} + 1)(1 - \eta)}, \quad (7.5.31)$$

where η is given by

$$\eta = \frac{q'_{ID} + 1}{q'_{IS} + 1}. \quad (7.5.32)$$

One should note that the term in square brackets in (7.5.31) is the normalized inversion charge in the channel of the device if $\zeta \rightarrow 0$.

Calculating $dq'_I/d\xi$ from (7.5.4) and substituting the result into (7.5.29) we find that

$$\frac{\overline{I_{d, heat}^2}}{4kTG_0\Delta f} = -2\zeta \frac{\int_{source}^{drain} \frac{q'_I (q'_I + 1)}{q'_I + 1 - \zeta i_D/2} dq'_I}{[1 + \zeta (q'_{IS} - q'_{ID})]^2}, \quad (7.5.33)$$

which integrated along the channel gives

$$\begin{aligned} \frac{\overline{I_{d, heat}^2}}{4kTG_0\Delta f} = & \frac{\zeta}{[1 + \zeta (q'_{IS} - q'_{ID})]^2} \left\{ (q'_{IS} - q'_{ID})(q'_{IS} + q'_{ID} + \zeta i_D) \right. \\ & \left. - \zeta i_D \left(1 - \frac{\zeta i_D}{2} \right) \ln \left(\frac{q'_{IS} + 1 - \frac{\zeta i_D}{2}}{q'_{ID} + 1 - \frac{\zeta i_D}{2}} \right) \right\}. \end{aligned} \quad (7.5.34)$$

In the linear region, for low drain-source voltages, the inversion charge along the channel is almost constant and, therefore, the transistor noise is similar to that of a long channel device.

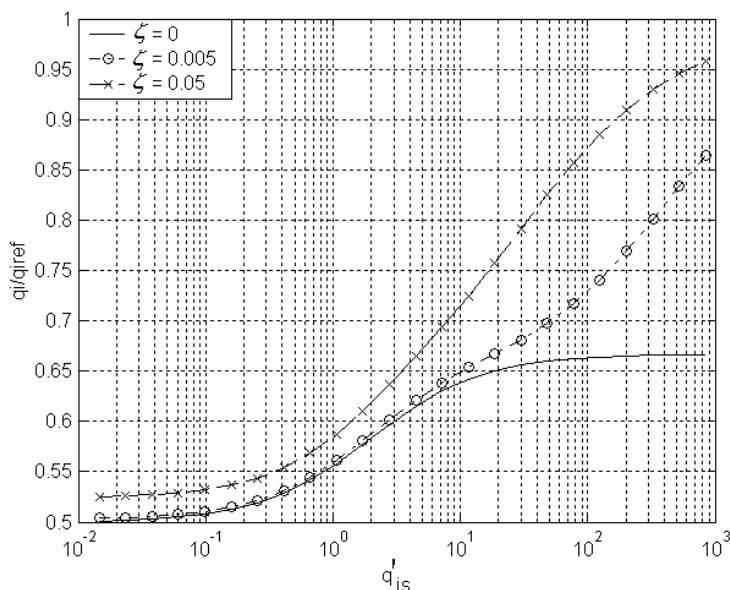


Fig. 7.18 Total channel inversion charge in saturation (normalized to the inversion charge for $V_{DS} = 0$) vs. inversion charge density at source for short-channel coefficient equal to 0 ($L \rightarrow \infty$), 0.005 ($L \approx 1 \mu\text{m}$) and 0.05 ($L \approx 0.1 \mu\text{m}$).

In Fig. 7.18 the total transistor charge is shown in the case of a transistor in saturation, while Fig. 7.19 displays the dependence of γ_{sat} , the “white noise gamma factor,” on both the inversion level at source and the short-channel factor. γ_{sat} for a saturated MOS transistor is defined as

$$\frac{\overline{I_d^2}}{4kT\Delta f} = \gamma_{sat} g_{d0}, \quad (7.5.35)$$

where g_{d0} is the MOSFET output conductance at $V_{DS}=0$. It is worth noting that, except for long-channel devices and not very high inversion levels, the usual expression for the channel noise based on the total charge does not hold.

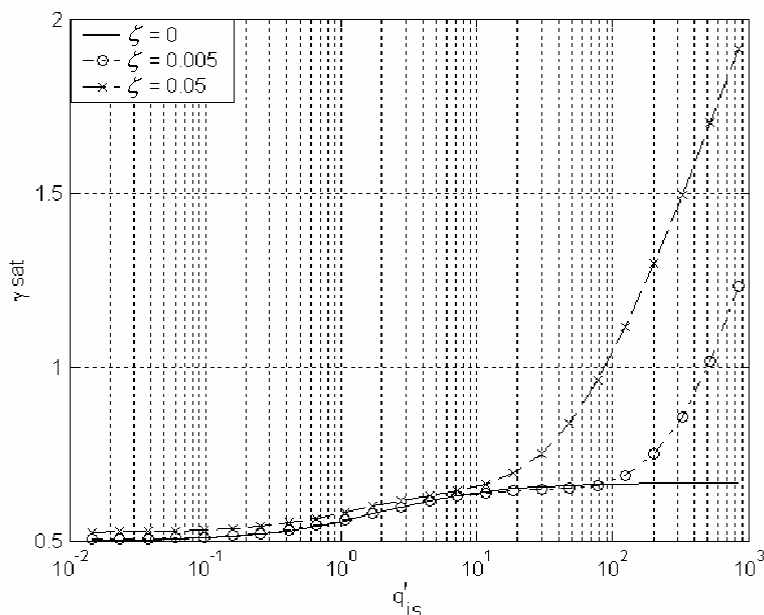


Fig. 7.19 Gamma factor vs. inversion charge density at source for short-channel coefficients equal to 0 ($L \rightarrow \infty$), 0.005 ($L \approx 1 \mu\text{m}$) and 0.05 ($L \approx 0.1 \mu\text{m}$).

Bibliography

- [1] C. T. Sah, "Theory and experiments on the $1/f$ surface noise of MOS insulated-gate field-effect transistors," *IEEE Trans. Electron Devices*, vol. 11, no. 11, p. 534, Nov. 1964.
- [2] A. G. Jordan and N. A. Jordan, "Theory of noise in metal oxide semiconductor devices," *IEEE Trans. Electron Devices*, vol. 12, no. 3, pp. 148-156, Mar. 1965.
- [3] H. K. J. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid-State Electron.*, vol. 7, no. 4, pp. 423-430, Apr. 1964.
- [4] C. T. Sah, "Characteristics of the metal-oxide-semiconductor transistor," *IEEE Trans. Electron Devices*, vol. 11, no. 7, pp. 324-345, July 1964.
- [5] A. van der Ziel, "Theory of noise in field effect transistors," *IEEE Trans. Electron Devices*, vol. 9, no. 11, p. 507, Nov. 1962.
- [6] C. T. Sah, "Theory of low-frequency generation noise in junction-gate field effect transistors," *Proc. IEEE*, vol. 52, no. 7, pp. 795-814, July 1964.
- [7] Albert Einstein, *Investigations on the Theory of the Brownian Movement*, Dover Publications, New York, 1956.
- [8] H. Nyquist, "Thermal agitation of electric charge in conductors," *Phys. Rev.*, vol. 32, no. 1, pp. 110-113, July 1928.

- [9] J. R. Pierce, "Physical sources of noise," *Proc. IRE*, vol. 44, pp. 601-608, May 1956.
- [10] J. B. Johnson, "Thermal agitation of electricity in conductors," *Phys. Rev.*, vol. 32, no. 1, pp. 97-109, July 1928.
- [11] Paul R. Gray and Robert G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed., John Wiley & Sons, New York, 1993.
- [12] J. R. Ragazzini and S. S. L. Chang, "Noise and random processes," *Proc. IRE*, vol. 50, pp. 1146-1151, May 1962.
- [13] Aldert van der Ziel, *Noise in Solid State Devices and Circuits*, Wiley-Interscience, New York, 1986.
- [14] Alfredo Arnaud Maceira, *Very Large Time Constant Gm-C Filters*, Ph. D. Thesis, Universidad de la Republica, Uruguay, 2004. [Online]. Available: http://eel.ufsc.br/~lci/work_doct.html
- [15] W. Shockley, John A. Copeland, and R. P. James, "The impedance field method of noise calculation in active semiconductor devices," pp. 537-563 in *Quantum Theory of Atoms, Molecules and the Solid-State*, Academic Press, New York, 1966.
- [16] Jung-Suk Goo, *High-Frequency Noise in CMOS Low Noise Amplifiers*, PhD Thesis, Stanford, CA, 2001. [Online]. Available: <http://www-tcad.stanford.edu/tcad/pubs/theses/goo.pdf>).
- [17] Richard S. C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New York, 1970.
- [18] Alfredo Arnaud and Carlos Galup-Montoro, "A compact model for flicker noise in MOS transistors for analog circuit design," *IEEE Trans. Electron Devices*, vol. 50, no. 8, pp. 1815-1818, Aug. 2003.
- [19] F. N. Hooge, "1/f noise sources," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 1926-1935, Nov. 1994.
- [20] C. T. Sah, S. Y. Wu, and F. H. Hielscher, "The effects of fixed bulk charge on the thermal noise in metal-oxide-semiconductor transistors," *IEEE Trans. Electron Devices*, vol. 13, no. 4, pp. 410-414, Apr. 1966.
- [21] S. Christensson, I. Lundstrom, and C. Svensson, "Low frequency noise in MOS transistors-I Theory," *Solid-State Electron.*, vol. 11, no. 9, pp.797-812, Sep. 1968.
- [22] F. Berz, "Theory of low frequency noise in Si MOSTs," *Solid-State Electron.*, vol. 13, no. 5, pp. 631-647, May 1970.
- [23] Kwok K. Hung, Ping K. Ko, Chenming Hu, and Yiu C. Cheng., "A physics-based MOSFET noise model for circuit simulators," *IEEE Trans. Electron Devices*, vol. 37, no. 5, pp. 1323-1333, May 1990.
- [24] G. Reimbold, "Modified 1/f trapping noise theory and experiments in MOS transistors biased from weak to strong inversion-influence of interface states," *IEEE Trans. Electron Devices*, vol. 31, no. 9, pp. 1190-1198, Sep. 1984.
- [25] Carlos Galup-Montoro, Márcio C. Schneider, and Itamar J. B. Loss, "Series-parallel association of FETs for high gain and high frequency applications," *IEEE J. Solid-State Circuits*, vol. 29, no. 9, pp. 1094-1101, Sep. 1994.
- [26] Christian C. Enz and Eric A. Vittoz, "CMOS low-power analog circuit design," pp. 79-133, in *Emerging Technologies*, Ralph Cavin and Wentai Liu (eds.), IEEE Press, Piscataway, NJ, 1996.

- [27] Alfredo Arnaud and Carlos Galup-Montoro, "Consistent noise models for analysis and design of CMOS circuits," *IEEE Trans. Circuits Systems I*, vol. 51, no. 10, pp. 1909-1915, Oct. 2004.
- [28] Kwangseok Han, Hyungcheol Shin, and Kwyro Lee, "Analytical drain thermal noise current model valid for deep submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 51, no. 2, pp. 261-269, Feb. 2004.
- [29] Carlo Jacobini and Lino Reggiani, "Bulk hot-electron properties of cubic semiconductors," *Advances in Physics*, vol. 28, no. 4, pp. 493-553, 1979.
- [30] Mark Lundstrom, *Fundamentals of Carrier Transport*, 2nd ed., Cambridge University Press, Cambridge, UK, 2000.

Problems

7.1. Demonstrate that the PSD of the drain current noise of a long-channel MOSFET can be written as

$$S_{th} = 8qI_S \left[\frac{2}{3} (q'_{IS} + 1) \left(\frac{1 + \alpha + \alpha^2}{1 + \alpha} \right) - 1 \right] \quad (\text{P7.1.1})$$

where q'_{IS} is the inversion charge at the drain normalized to the pinch-off charge and $\alpha = (q'_{ID} + 1)/(q'_{IS} + 1)$. Use this expression to calculate the approximate thermal noise PSD in weak inversion and strong inversion, for both the linear and saturation regions.

7.2. The following formulas are used to represent the PSD of the MOSFET flicker noise:

$$S_{Id} = \frac{K_F I_D^{AF}}{C'_{ox} L^2} \frac{1}{f} \quad (\text{P7.2.1})$$

$$S_{Id} = \frac{K_F I_D^{AF}}{C'_{ox} WL} \frac{1}{f} \quad (\text{P7.2.2})$$

$$S_{Id} = \frac{K_F g_{mg}^2}{C'_{ox} WL} \frac{1}{f^{EF}} \quad (\text{P7.2.3})$$

for **noiselevel**=0, 1, and 2/3, respectively. Verify whether these formulas are consistent for the series and parallel association of transistors and specify the value of AF required for consistency. Also determine the

asymptotic value of the normalized PSD S_{Id}/I_D^2 for deep weak inversion (for a discussion on the consistency of the models, see [27]).

7.3. Show that, in weak inversion, the thermal noise of a MOSFET can be written as the sum of the shot noise components from the forward (I_F) and reverse (I_R) currents (a discussion on this subject can found in R. Sarpeshkar, T. Delbrück, and C. A. Mead, “White noise in MOS transistors and resistors”, IEEE Circuits & Devices Mag., vol. 9, no. 6, pp.23-29, Nov. 1993).

This page intentionally left blank

Chapter 8

High-Frequency Models

The dynamic MOSFET model we have presented so far assumes the transistor to be in quasi-static operation, *i.e.*, the charge densities at any position in the channel are assumed to depend on the instantaneous values of the terminal voltages only [1], [2], [3]. If, however, the rate of variation of the terminal voltages is high, the quasi-static approximation is no longer valid. In fact, the charge densities along the channel depend not only on the voltage values at a certain time but also on the history leading to the charge densities at that time [2]. The model used for fast varying signals is called the non-quasi-static model, which is the subject of this chapter.

8.1 Introduction

In chapter 5 we described the intrinsic quasi-static charge-conserving model. The quasi-static approximation assumes that the charge density at any position in the channel changes instantaneously with the applied voltages [1], [2], [4], *i.e.*, it assumes the transit time in the channel to be zero [4]. Since the channel transit time is not zero, analyses based on the quasi-static approximation introduce errors for rapidly changing terminal voltages due to the distributed nature of the MOSFET. Models that take into account the distributed nature of the transistor are described as non-quasi-static (NQS) models. The derivation of NQS models is not a simple task since it requires the solution of both transport and continuity equations. In addition, the transport equation is non-linear, making the problem even more difficult to solve.

In the sections that follow, we will describe the equations associated with the non-quasi-static model and make simplifying assumptions to solve the equations for the particular case of small signals applied to the device terminals. More complicated methods that include NQS effects in the transistor model are the subject of intense investigation but are beyond the scope of this text. The small-signal NQS model for any bias regime, that we show in this chapter, is quite similar to those derived in [1], [5], [6] for strong inversion. The main modification of our formulas is that we define a general saturation index for any operating condition, rather than particularizing the saturation index for strong inversion. In the last part of this chapter we derive expressions for both the power spectral density of the induced gate noise and the correlation coefficient between the induced gate noise and the channel thermal noise. The noise models we show in this chapter are similar to those presented in [7] and [8].

8.2 Non quasi-static operation of the MOSFET

The NQS operation of the MOS transistor is described by the continuity equation and by the transport equation. Neglecting recombination in the channel, the current-continuity equation is given by Eq. (5.1.3), rewritten below for convenience

$$\frac{\partial I(y,t)}{\partial y} = W \frac{\partial Q'_i(y,t)}{\partial t}. \quad (8.2.1)$$

The channel current $I(y,t)$ is assumed to flow in the direction of the negative y-axis. The Pao-Sah current transport equation is given by Eq. (5.1.4), rewritten below

$$I(y,t) = -\mu W Q'_i(y,t) \frac{\partial V_c(y,t)}{\partial y} \quad (8.2.2)$$

where $V_c(y,t)$ in (8.2.2) is the channel potential. Now, using UCCM in differential form, (3.5.1), the right-hand side of (8.2.2) can be written in terms of the inversion charge density, yielding

$$I(y, t) = -\frac{1}{2} \frac{\mu W}{nC'_{ox}} \frac{\partial \left(Q'_l(y, t) - nC'_{ox} \phi_t \right)^2}{\partial y}. \quad (8.2.3)$$

The quasi-static approximation assumes that the terminal voltages vary slowly enough for the charge stored to follow instantaneously the voltage variations [1], [9]; as a result, charge variations along the channel are so slow that the right-hand side of (8.2.1) can be made equal to zero, which implies that the current does not change along the channel. It has been pointed out in [1] that, in strong inversion saturation, the quasi-static approximation is valid only if the device transit time is of the order of 20 times shorter than the rise or fall times of the waveforms involved.

The transient performance of the MOS transistor in high-speed circuits can be evaluated by solving simultaneously equations (8.2.1) and (8.2.3) and using appropriate boundary conditions. Closed-form solutions are difficult to obtain for the transient performance; therefore, numerical methods are commonly employed to calculate the time performance of high-speed circuits. Since in this text the analysis of the MOSFET performance will be limited to small-signal analysis, the interested reader is referred to [10], [11] for more details about large-signal analysis of MOS circuits with rapidly varying signals.

8.3 Non quasi-static small-signal model

Differentiating (8.2.3) wrt y and equating the result to (8.2.1) gives

$$\frac{\partial Q'_l(y, t)}{\partial t} = -\frac{1}{2} \frac{\mu}{nC'_{ox}} \frac{\partial^2 \left(Q'_l(y, t) - nC'_{ox} \phi_t \right)^2}{\partial y^2}. \quad (8.3.1)$$

Before deriving the small-signal NQS model, we calculate the quasi-static solution of (8.3.1) by making the time derivative equal to zero. The double integration of (8.3.1) results in

$$\left(\frac{Q'_l(y)}{-nC'_{ox} \phi_t} + 1 \right)^2 = \left(\frac{Q'_l(0)}{-nC'_{ox} \phi_t} + 1 \right)^2 \left(1 - \frac{y}{L} \right) + \frac{y}{L} \left(\frac{Q'_l(L)}{-nC'_{ox} \phi_t} + 1 \right)^2. \quad (8.3.2)$$

To derive the small-signal parameters of the NQS model, let us assume the external voltages to be sine waves of so small a magnitude that the charge and current variations along the channel are also sinusoidal and can be computed using small-signal analysis. The inversion charge density and current along the channel can then be expressed as

$$Q'_l(y, t) = Q'_l(y) + Q'_i(y, \omega) e^{j\omega t} \quad (8.3.3)$$

$$I(y, t) = I_D + I_i(y, \omega) e^{j\omega t}. \quad (8.3.4)$$

The substitution of (8.3.3) and (8.3.4) into (8.2.3) yields, for small signals

$$I_D + I_i(y, \omega) e^{j\omega t} = -\frac{1}{2} \frac{\mu W}{nC'_{ox}} \frac{\partial [Q'^2_{lt}(y) + 2Q'_{lt}(y)Q'_i(y, \omega) e^{j\omega t}]}{\partial y} \quad (8.3.5)$$

where $Q'_{lt}(y) = Q'_l(y) - nC'_{ox}\phi_t$ is the shifted charge. In (8.3.5), one can readily identify the ac magnitude of the current as

$$I_i(y, \omega) = -\frac{\mu W}{nC'_{ox}} \frac{\partial [Q'_{lt}(y)Q'_i(y, \omega)]}{\partial y}. \quad (8.3.6)$$

The continuity equation given by (8.2.1) along with (8.3.3) and (8.3.4) yields

$$\frac{\partial I_i(y, \omega)}{\partial y} = j\omega W Q'_i(y, \omega). \quad (8.3.7)$$

Differentiating (8.3.6) with respect to y and equating the result to (8.3.7), we find that

$$\frac{\partial^2 [Q'_{lt}(y)Q'_i(y, \omega)]}{\partial y^2} + \frac{j\omega nC'_{ox}}{\mu} Q'_i(y, \omega) = 0. \quad (8.3.8)$$

The solution of (8.3.8) is given in terms of Bessel functions of the order $1/3$ and $-1/3$ [12], [13] with complex arguments, but is not of much use, especially if one is interested in compact models. Thus, a simpler

model to describe the small-signal NQS characteristics is highly desirable.

In the paragraphs that follow, we will derive a set of equations written as rational functions of the frequency, which are relatively simple and highly useful for determining high-frequency characteristics of MOS transistors. The method to be followed, presented in [14], allowed the determination of closed-form expressions for the small-signal parameters of a saturated MOSFET in strong inversion by iterative integration of expressions (8.3.6) and (8.3.7). The method proposed in [14] has also been applied by the authors of [1], [5], [6] for all operating regimes, but the results are not easily interpreted except for strong inversion. The reason for this difficulty is that at the time of these publications, an analytic compact dc model of the MOS transistor including all operating regions was not available. In the second half of the 1990's, three publications [15], [16], [17] presented a compact small-signal NQS for the MOSFET valid in all operating regions. In the following, we reproduce the main results of these papers.

The integration of (8.3.6) and (8.3.7) from an arbitrary position in the channel to the drain gives, respectively,

$$\int_y^L I_i(\hat{y}, \omega) d\hat{y} = \mu\phi_t W \left[\frac{Q'_t(L)}{Q'_{IP}} Q'_i(L, \omega) - \frac{Q'_t(y)}{Q'_{IP}} Q'_i(y, \omega) \right] \quad (8.3.9)$$

and

$$I_i(y, \omega) = I_i(L, \omega) - j\omega W \int_y^L Q'_i(\hat{y}, \omega) d\hat{y}. \quad (8.3.10)$$

The method used in [14] to find solutions to $I_i(y, \omega)$ and $Q'_i(y, \omega)$ consists of the iterative integration of (8.3.9) and (8.3.10). To illustrate the method, we start with the zero order (quasi-static) approximation ($\omega=0$). In this case, we write (8.3.10) as

$$I_{i0}(y, \omega) = I_i(L, \omega). \quad (8.3.11)$$

The subscript zero refers to the zero order approximation. In the next step, the current from (8.3.11) is substituted in (8.3.9) and integrated from a generic position y along the channel to the drain end ($y=L$). The integration is calculated after making the following change of variable

$$dy = \frac{-\mu W}{nC'_{ox} I_D} Q'_{it} dQ'_{it}. \quad (8.3.12)$$

The integration of expression (8.3.9) allows the determination of $Q'_{i0}(y, \omega)$, the zero order approximation of the ac magnitude of the normalized inversion charge along the channel. To find $I_{i1}(y, \omega)$, the first order approximation of the current along the channel, the charge density $Q'_{i0}(y, \omega)$ is substituted in (8.3.10) and integrated. In the next step, the first order approximation $Q'_{i1}(y, \omega)$ is found from (8.3.9). For more details about the successive approximations of the iterative procedure, the reader is referred to [6], [15].

Figure 8.1 illustrates the currents of the MOSFET. The channel current I_i is assumed to flow from drain to source. The gate and bulk currents are assumed to enter into the device.

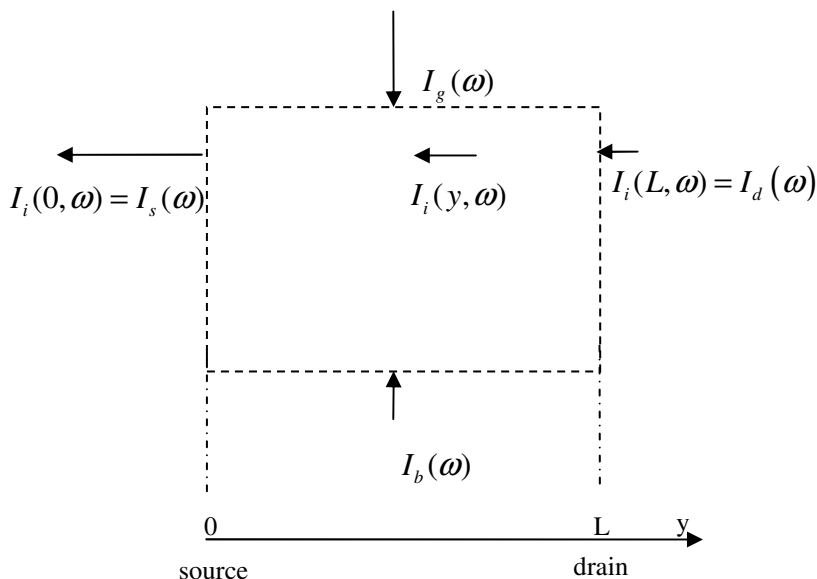


Fig. 8.1 Illustration of a closed surface enclosing the MOSFET and the terminal currents.

The drain current $I_i(L, \omega)$ can be determined by substituting the preceding approximation of the current and integrating (8.3.9) from source to drain. The result obtained after substituting the first order approximation of the current in (8.3.9) and integrating from source to drain is

$$I_i(L, \omega) = \frac{\mu W Q'_F}{nC'_{ox} L D(\omega)} [Q'_i(0, \omega) - \alpha Q'_i(L, \omega)] + j\omega W L \frac{Q'_i(L, \omega)}{D(\omega)} P_\alpha \quad (8.3.13)$$

where

$$D(\omega) = 1 - \frac{j\omega n C'_{ox} L^2}{\mu Q'_F} \frac{4}{15} \frac{1 + 3\alpha + \alpha^2}{(1 + \alpha)^3} \quad (8.3.14)$$

$$P_\alpha = \frac{2}{3} \alpha \frac{(2 + \alpha)}{(1 + \alpha)^2} \quad (8.3.15)$$

$$Q'_F = Q'_i(y = 0), \quad (8.3.16)$$

$$Q'_R = Q'_i(y = L), \quad (8.3.17)$$

and

$$\alpha = \frac{Q'_i(L)}{Q'_i(0)} = \frac{Q'_R}{Q'_F}. \quad (8.3.18)$$

(8.3.13) is a generalized expression for the ac drain current of a long-channel device in any operating region. It includes weak, moderate, and strong inversion, through the inversion level at the source, and also describes the ac current from the linear to the saturation region, through parameter α . In addition, (8.3.13) is similar to the formulas presented in [6], which were derived for strong inversion. A close comparison between (8.3.13) and the result in [6] reveals that the same expression used in [6] for calculating the ac drain current in strong inversion can be

used, by proper interpretation of dc parameters, to determine the ac current for any inversion level.

The source current can be determined from the same expression as the drain current by using the symmetry property of the MOSFET, *i. e.*, the source current equals the drain current if drain and source are interchanged. Using symmetry and expression (8.3.13) we find that

$$I_i(0, \omega) = \frac{\mu W Q'_F}{nC'_{ox} LD(\omega)} \left[Q'_i(0, \omega) - \alpha Q'_i(L, \omega) \right] - j\omega WL \frac{Q'_i(0, \omega)}{D(\omega)} P_{1/\alpha} \quad (8.3.19)$$

To derive (8.3.19), we substituted Q'_R for Q'_F , $1/\alpha$ for α , $Q'_i(0, \omega)$ for $Q'_i(L, \omega)$, and $Q'_i(L, \omega)$ for $Q'_i(0, \omega)$. In addition, we changed the current sign to keep the current flow in the negative y-direction.

In order to calculate the current associated with the bulk charge variation, we rewrite here the approximate relationship between the bulk charge density and the inversion charge density

$$Q'_B = -\frac{n-1}{n} Q'_I + Q'_{Ba} \quad (8.3.20)$$

where,

$$Q'_{Ba} = -\gamma C'_{ox} \left(\sqrt{V_{GB} - V_{FB} - \phi_t + \gamma^2/4} - \gamma/2 \right) \quad (8.3.21)$$

is the depletion charge density for $Q'_I = 0$.

For the derivation of the bulk current, we have assumed that expression (8.3.20) is valid for any instant, *i. e.*, the bulk charge density and the inversion charge density track each other instantaneously. Moreover, the charge density Q'_{Ba} is assumed to respond instantaneously to either the gate voltage or the bulk voltage.

The substrate current, which is assumed to enter the device, is given by

$$I_B(t) = \frac{dQ_B(t)}{dt} = W \frac{d}{dt} \int_0^L Q'_B(y, t) dy \quad (8.3.22)$$

or, using (8.3.20), the substrate current becomes

$$I_B(t) = W \frac{d}{dt} \int_0^L \left[-\frac{n-1}{n} Q'_l(y, t) + Q'_{Ba}(t) \right] dy. \quad (8.3.23)$$

Since Q'_{Ba} does not depend on y , (8.3.23) can be rewritten as

$$I_B(t) = W \frac{d}{dt} \int_0^L -\frac{n-1}{n} Q'_l(y, t) dy + WL \frac{d}{dt} Q'_{Ba}(t). \quad (8.3.24)$$

We now write the expression for the substrate current in the frequency domain as

$$I_B(\omega) = I_b(\omega) e^{j\omega t}, \quad (8.3.25)$$

and the depletion charge density for deep weak inversion as

$$Q'_{Ba}(t) = Q'_{Ba} + Q'_{ba}(\omega) e^{j\omega t}, \quad (8.3.26)$$

Expression (8.3.24) written in the frequency domain then becomes

$$I_b(\omega) = -j\omega WL \left[\frac{n-1}{n} \frac{1}{L} \int_0^L Q'_l(y, \omega) dy - Q'_{ba}(\omega) \right]. \quad (8.3.27)$$

In order to solve (8.3.27), we have used $Q'_{il}(y, \omega)$, the first order approximation of the ac magnitude of the inversion charge density. As stated before, the depletion charge in deep weak inversion, $Q'_{Ba}(t)$, is assumed to vary instantaneously with the gate or bulk voltage. The ac magnitude of the normalized substrate current is

$$\begin{aligned} I_b(\omega) = & -j\omega WL \frac{n-1}{n} \left[Q'_l(L, \omega) M_\alpha(\omega) \right. \\ & \left. + \frac{2L}{\mu W \phi_t} I_d(\omega) N_\alpha \right] + j\omega WL Q'_{ba}(\omega) \end{aligned} \quad (8.3.28)$$

with

$$M_\alpha(\omega) = \frac{2\alpha}{1+\alpha} - \frac{2}{3} \frac{j\omega n C'_{ox} L^2}{\mu Q'_F} \frac{\alpha}{(1+\alpha)^2} \quad (8.3.29)$$

and

$$N_{\alpha} = \frac{nC'_{ox}\phi_t}{Q'_F} \left[\frac{2}{3} \frac{1+2\alpha}{(1+\alpha)^2} \right]. \quad (8.3.30)$$

Finally, to calculate the gate current, we apply the charge conservation principle

$$\frac{dQ_I(t)}{dt} + \frac{dQ_B(t)}{dt} + \frac{dQ_G(t)}{dt} = \frac{dQ_I(t)}{dt} + I_B(t) + I_G(t) = 0. \quad (8.3.31)$$

Now, inserting the value of $I_B(t)$ from (8.3.23) into (8.3.31), we find that

$$I_G(t) = -W \frac{d}{dt} \left[\int_0^L \frac{1}{n} Q'_i(y, t) dy + LQ'_{Ba}(t) \right]. \quad (8.3.32)$$

We now write the expression for the normalized substrate current in the frequency domain as

$$I_G(t) = I_g(\omega) e^{j\omega t}. \quad (8.3.33)$$

Expression (8.3.32) written in the frequency domain then becomes

$$I_g(\omega) = -j\omega W \left[\frac{1}{n} \int_0^L Q'_i(y, \omega) dy + LQ'_{ba}(\omega) \right]. \quad (8.3.34)$$

The ac magnitude of the gate current is

$$I_g(\omega) = j\omega WL \frac{1}{n} \left[Q'_i(L, \omega) M_{\alpha}(\omega) + \frac{2LI_d(\omega)}{\mu W \phi_t} N_{\alpha} \right] - j\omega WL Q'_{ba}(\omega). \quad (8.3.35)$$

Having determined the source, drain, bulk, and gate currents, we are now in a position to determine the small-signal parameters of the MOSFET. Before doing this, we must calculate the variation of the inversion charge densities at source and drain in terms of the variations of the terminal voltages. For this, we rewrite UCCM in its approximate form for the source and drain ends.

$$\frac{V_G(t) - V_B(t)}{n\phi_t} - \frac{V_S(t) - V_B(t)}{\phi_t} = \frac{Q'_i(0, t)}{Q'_{IP}} - 1 + \ln \frac{Q'_i(0, t)}{Q'_{IP}} \quad (8.3.36)$$

$$\frac{V_G(t) - V_B(t)}{n\phi_t} - \frac{V_D(t) - V_B(t)}{\phi_t} = \frac{Q'_i(L, t)}{Q'_{IP}} - 1 + \ln \frac{Q'_i(L, t)}{Q'_{IP}}. \quad (8.3.37)$$

Let us assume that the voltages are exponential signals, *i.e.*

$$V_M(t) = V_M + V_m e^{j\omega t} \quad (8.3.38)$$

where $M(m) = S(s)$, $D(d)$, $G(g)$, or $B(b)$ is any of the MOSFET terminals. Now, using small-signal analysis for the charges ($Q'_i(0) \ll |Q'_i(0)|$ and $Q'_i(L) \ll |Q'_i(L)|$) we find the relationship between the ac magnitude of the inversion charge density at source (drain) and the ac magnitudes of the voltages V_g , $V_{s(d)}$, V_b

$$Q'_i(0) = \left[\frac{Q'_i(0)}{1 + Q'_i(0)/Q'_{IP}} \right] \left[\frac{V_g}{n\phi_t} - \frac{V_s}{\phi_t} + \frac{n-1}{n\phi_t} V_b \right] \quad (8.3.39)$$

$$Q'_i(L) = \left[\frac{Q'_i(L)}{1 + Q'_i(L)/Q'_{IP}} \right] \left[\frac{V_g}{n\phi_t} - \frac{V_d}{\phi_t} + \frac{n-1}{n\phi_t} V_b \right]. \quad (8.3.40)$$

The ac magnitude of the depletion charge Q'_{ba} depends on V_g and V_b in accordance with

$$Q'_{ba} = -\frac{n-1}{n} C'_{ox} (V_g - V_b). \quad (8.3.41)$$

Expression (8.3.41) was derived from (8.3.21) assuming a small variation of the depletion charge as compared with the quiescent depletion charge. The results we have derived in this section will be used to determine the high-frequency y-parameter model of the MOS transistor.

8.4 The y-parameter model

The y parameters of the MOS transistor are defined according to the admittance matrix representation [1] given below

$$\begin{pmatrix} I_d \\ I_g \\ I_b \\ I_s \end{pmatrix} = \begin{pmatrix} Y_{dd} & Y_{dg} & Y_{db} & Y_{ds} \\ Y_{gd} & Y_{gg} & Y_{gb} & Y_{gs} \\ Y_{bd} & Y_{bg} & Y_{bb} & Y_{bs} \\ Y_{sd} & Y_{sg} & Y_{sb} & Y_{ss} \end{pmatrix} \begin{pmatrix} V_d \\ V_g \\ V_b \\ V_s \end{pmatrix}. \quad (8.4.1)$$

The sixteen parameters in (8.4.1) are not independent. In fact, one can select nine independent parameters only. As for the capacitive coefficients, charge conservation and the fact that only three voltage differences out of four can be chosen independently, implies that a maximum of nine admittances only are independent.

To determine the Y parameters we use the results of the previous section. Expressions (8.3.13), (8.3.19), (8.3.28), and (8.3.35) give the terminal currents in terms of the inversion charge densities at the drain and source ends of the channel, and the depletion charge density for deep weak inversion. The dependence of the charge densities on the terminal voltages are given in expressions (8.3.39), (8.3.40), and (8.3.41). This set of seven expressions is enough to determine the admittance parameters. The following simplifications will be carried out in our calculations: the derivative of the slope factor with respect to the gate voltage will be considered negligible while the rational functions representing the NQS effects will have the frequency terms of order equal or higher than two truncated.

By applying the definitions of admittances together with expressions (8.3.13), (8.3.19), (8.3.28), and (8.3.35) for the terminal currents in terms of terminal charges, and (8.3.39), (8.3.40), and (8.3.41), which give the terminal charges as functions of the terminal voltages, we obtained the set of nine independent (trans)admittances presented in Table 8.1.

The expressions in Table 8.1 are similar to the corresponding ones derived in [5]. However, the NQS parameters shown here are expressed in terms of the inversion charge densities at source and drain while the parameters in [5] are functions of the surface potentials at source and drain. The expressions in Table 8.1, with the exception of τ_4 , are similar to those given in [1]. However, Table 8.1 is a general description of the NQS model for any operating condition, while the results in [1] are presented for strong inversion only.

Table 8.1 Admittances of the NQS MOSFET model

$-y_{gs} = j\omega C_{gs} \frac{1+j\omega\tau_2}{1+j\omega\tau_1}, \quad y_{bs} = (n-1)y_{gs}$
$-y_{gd} = j\omega C_{gd} \frac{1+j\omega\tau_3}{1+j\omega\tau_1}, \quad y_{bd} = (n-1)y_{gd}$
$-y_{gb} = -y_{bg} = j\omega C_{gb} \frac{1+j\omega\tau_4}{1+j\omega\tau_1}$
$-y_{ds} = \frac{g_{ms}}{1+j\omega\tau_1}, \quad -y_{sd} = \frac{g_{md}}{1+j\omega\tau_1}, \quad y_{dg} = \frac{g_{mg}}{1+j\omega\tau_1} + y_{gd}$
$\tau_1 = \frac{\tau}{1+q'_{IS}} \frac{4}{15} \frac{1+3\alpha+\alpha^2}{(1+\alpha)^3}$
$\tau_2 = \frac{\tau}{1+q'_{IS}} \frac{1}{15} \frac{2+8\alpha+5\alpha^2}{(1+\alpha)^2(1+2\alpha)}$
$\tau_3 = \frac{\tau}{1+q'_{IS}} \frac{1}{15} \frac{5+8\alpha+2\alpha^2}{(1+\alpha)^2(2+\alpha)} \quad \tau_4 = \frac{C_{ox}\tau_1 - C_{gs}\tau_2 - C_{gd}\tau_3}{C_{ox} - C_{gs} - C_{gd}}$
$\tau = \frac{L^2}{\mu\phi_t}; \quad \alpha = \frac{1+q'_{ID}}{1+q'_{IS}}; \quad q'_{IS(D)} = \frac{Q'_{IS(D)}}{-nC'_{ox}\phi_t}$

From the expressions in Table 8.1, it can be readily verified that:

$$C_{sd(ds)} = -g_{md(s)}\tau_1; \quad C_m = C_{dg} - C_{gd} = C_{gs} - C_{sg} = g_{mg}\tau_1 \quad (8.4.2)$$

To derive the first equality in (8.4.2), let us assume $\omega\tau_1 \ll 1$, and use the approximation $1/(1+j\omega\tau_1) \cong 1-j\omega\tau_1$ for y_{sd} , y_{ds} , and y_{gd} . Comparing the results thus obtained with the quasi-static model for y_{sd} , y_{ds} , and y_{gd} we find the equalities in (8.4.2).

Now let us interpret the time constant τ_1 . In the first order analysis of the NQS model, the frequency ω_{crit}^1 of the dominant pole of the y-parameters is equal to $1/\tau_1$. This value of ω_{crit} sets an upper limit for the validity of the quasi-static model, which is an acceptable representation of the MOSFET behavior for frequencies well below ω_{crit} .

$\tau_{2(3)}$ is always of the order of $\tau_1/2$, which allows one to represent the first order small-signal NQS model as the simplified schematic shown in Fig. 8.2 as long as we neglect the effects of τ_4 .

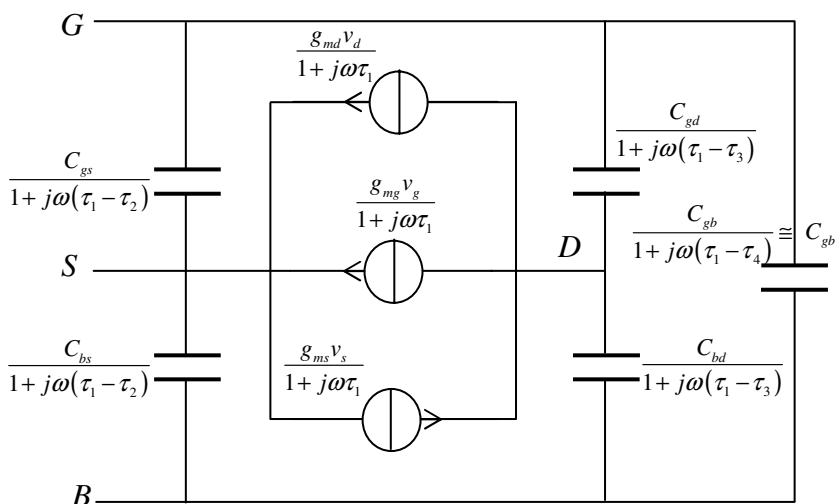


Fig. 8.2 Simplified high-frequency MOSFET model.

Now let us determine the frequency limit of validity of the simplified high-frequency MOSFET model in Fig. 8.2 in the case of more practical interest, *i.e.*, when the MOSFET is in saturation. For $i_r \rightarrow 0$, the first-order time constant τ_1 in Table 8.1 can be written in terms of the forward current as

$$\tau_1 = \tau \frac{4}{15} \frac{2 + i_f + 3\sqrt{1 + i_f}}{(\sqrt{1 + i_f} + 1)^3}. \quad (8.4.3)$$

¹ The notation ω_{crit} to characterize the NQS model has been used by the authors of [7] and [8].

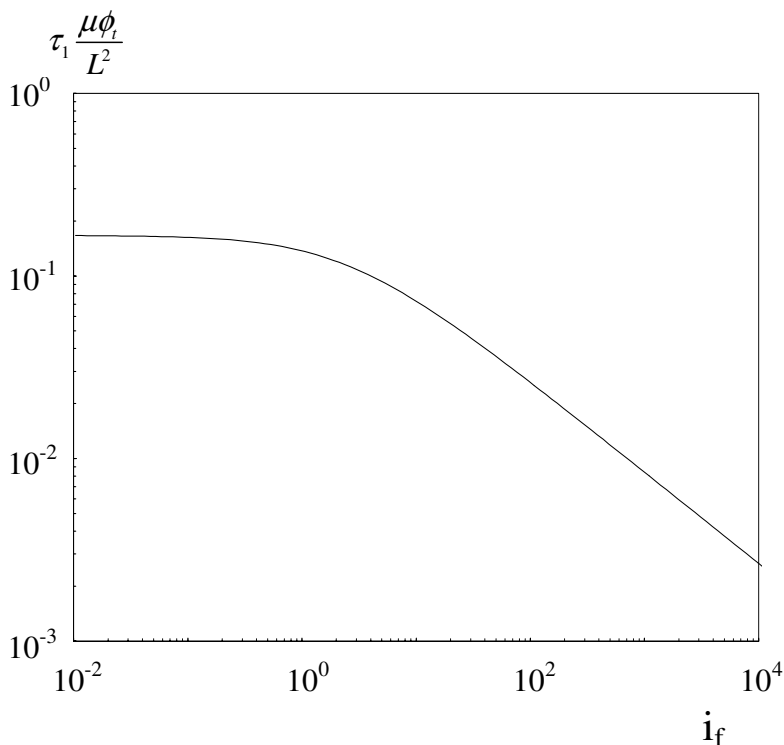


Fig. 8.3 Normalized time constant of high-frequency model in saturation.

Fig. 8.3 illustrates the variation of time constant τ_1 in saturation, with the forward normalized current. Assuming for the moment that the transistor will be used up to the intrinsic cutoff frequency, which is proportional to $\sqrt{1+i_f}-1$, the non-quasi-static correction of the transadmittances is only important for moderate and strong inversion. In moderate and strong inversion, the applicability of the quasi-static model should be restricted to frequencies up to one-third of the intrinsic cutoff frequency [1]. In weak inversion, the time constant τ_1 is a constant of the order of $\pi/6$, while the intrinsic cutoff frequency is proportional to the inversion level. Therefore, in weak inversion, the quasi-static model predicts dynamic operation with satisfactory accuracy at frequencies up to the inverse of the transit time, $L^2/(2\mu\phi_t)$.

Finally, as demonstrated in reference [17], and in [1] for strong inversion, the dynamic NQS model where $-y_{gb}$ is approximated by the constant capacitance C_{gb} represents fairly well the first-order high-frequency behavior of the MOS transistor for any operating regime.

As presented in [17], τ_4 is always greater than τ_1 , but in weak inversion τ_4 and τ_1 are about the same. Consequently, in weak inversion, the non-quasi-static effects in $C_{gb(bg)}$ are canceled out.

8.5 Channel segmentation

One method to model the transistor for fast signal variations is to consider it as a distributed transistor consisting of N equal channel segments in series, as represented in Fig. 8.4 [1], [3], [4], [18], [19].

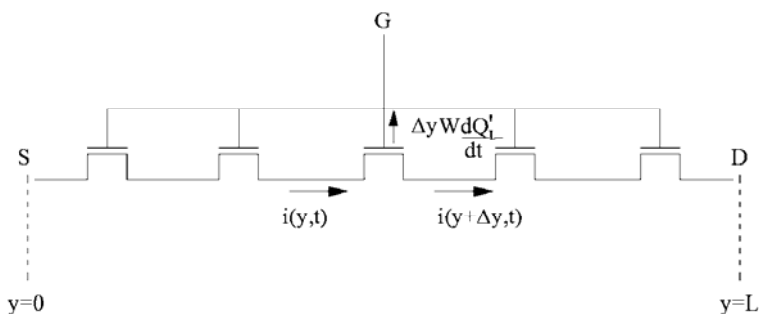


Fig. 8.4 Representation of the MOSFET as a series combination of segmented elements.

With the exception of the extreme segments, each of the segments consists only of intrinsic parts. The advantage of this method is that each of the segments can be modeled quasi-statically provided that each section is “short enough”. The difficult with the method is to define what “short enough” means, *i.e.*, what the number of segments should be in order to represent adequately the MOS transistor [18]. The decision regarding the number of segments is based on successive simulations of the device with an increasing number of segments and comparison of the successive results of the simulations. When the simulation differences

between successive steps are within an acceptable error, the number of adequate segments has been reached.

8.6 Induced gate noise

The origin of the gate noise of a MOSFET is the thermal noise of the conducting channel that is capacitively coupled to the gate. Due to the distributed thermal noise in the channel, noise voltages (or charge density fluctuations) will develop along the channel [20]. Because of the capacitive coupling between channel and gate, a noise current will flow through the transistor gate. This noise is referred to as induced gate noise. To understand how to derive the induced gate noise, let us consider the MOSFET channel represented as an active transmission line at high frequencies, as described in [21] and illustrated in Fig. 8.5. The resultant expressions of [21], however, are quite difficult to understand and to apply to practical problems. Therefore, we will limit our analysis of the induced gate noise to moderately high frequencies, to be quantified later, where the quasi-static transistor model is valid. To simplify the analysis and arrive at manageable results, we will assume that the gate current is substantially smaller than the channel current. This simplification is equivalent to saying that the channel current does not change along the channel. Each channel element gives rise to channel noise which is represented by a current source $i_{\Delta A}$ in parallel with the channel element. Fig. 8.5 shows just one such a current source. The current Δi_d that flows in the channel as a result of the elemental noise disturbance causes a voltage drop along the channel and a corresponding fluctuation of the inversion charge density. This voltage drop is coupled to the gate through the gate oxide capacitance, giving rise to a gate current. The superposition of all the small channel elements gives rise to the total induced gate noise.

To derive the equations that characterize the gate noise, we apply a procedure developed by van der Ziel [20], [22], but we use the fluctuation in inversion charge density rather than in voltage to characterize noise. The starting point of the analysis is the PSD of the

diffusion noise associated with an infinitesimally small segment $\Delta A = W \Delta y$ of semiconductor, which is given [22] by

$$\overline{i_{\Delta A}^2} = -4qD_n Q'_l(y) \frac{W}{\Delta y} \quad (8.6.1)$$

for an infinitesimally small thermal noise source localized between y and $y + \Delta y$. Using the current division property of a MOSFET, as previously done in the derivation of flicker noise in Section 7.2 using the three-transistor model, the drain current generated by $i_{\Delta A}$ is given by

$$\Delta i_d = \frac{\Delta y}{L} i_{\Delta A} \quad (8.6.2)$$

under the assumption of quasi-static operation. Equation (8.6.2) is valid up to frequencies where the current that deviates to the gate is much smaller than the channel current.

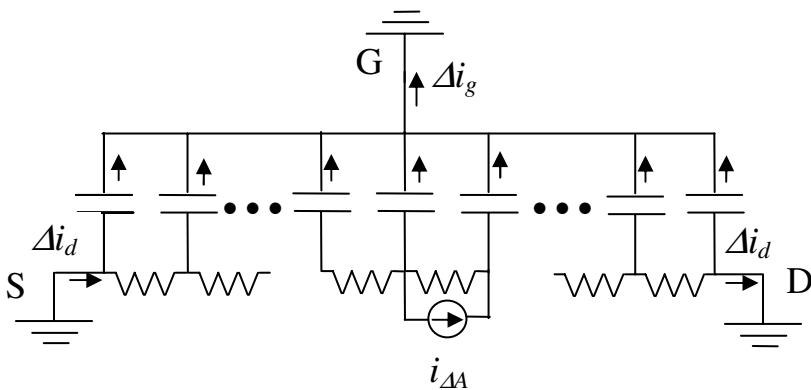


Fig. 8.5 Distributed ac MOSFET model used for calculation of the induced gate noise.

The relationship between the channel current and the fluctuation of the inversion charge density along the channel is found from expression (8.3.6), which is rewritten as

$$\Delta i_d = -\frac{\mu W}{nC'_{ox}} \frac{\partial [Q'_l(y) \Delta Q'_l(y)]}{\partial y} \quad (8.6.3)$$

The integration of (8.6.3) from the source to a particular position $y < y_0$ along the channel results in

$$\int_0^y \Delta i_d d\hat{y} = \Delta i_d (y - 0) = -\frac{\mu W}{nC'_{ox}} Q'_t \Delta Q'_i(y). \quad (8.6.4)$$

From (8.6.4), the inversion charge density for $0 < y < y_0$ is

$$\Delta Q'_i(y) = -\frac{nC'_{ox}}{\mu W Q'_t} y \Delta i_d. \quad (8.6.5)$$

By the same token, the integration of (8.6.3) from a generic position y such that $y_0 < y < L$ to the drain results in

$$\int_y^L \Delta i_d d\hat{y} = \Delta i_d (L - y) = \frac{\mu W}{nC'_{ox}} Q'_t \Delta Q'_i(y). \quad (8.6.6)$$

The noise charge density is then written as

$$\Delta Q'_i(y) = \frac{nC'_{ox}}{\mu W Q'_t} (L - y) \Delta i_d \quad (8.6.7)$$

for $y_0 < y < L$.

A local disturbance $\Delta Q'_{i0}$ in the charge density at y_0 affects the inversion charge densities along the channel in such a way as to ensure that the current is continuous along the channel. The variation $\Delta Q'_{i0}$ in the inversion charge density due to local thermal fluctuation is represented by the jump in Fig. 8.6. Note that the jump $\Delta Q'_{i0}$ can be calculated from the difference between (8.6.7) and (8.6.5) calculated for $y = y_0$, which yields

$$\Delta Q'_{i0} = \frac{nC'_{ox} L}{\mu W} \frac{\Delta i_d}{Q'_t(y_0)}. \quad (8.6.8)$$

Now, to derive the gate current that originates from the channel noise, we rewrite the gate current in terms of the inversion charge density

$$I_G(t) = -W \frac{d}{dt} \left[\int_0^L \frac{1}{n} Q'_i(y, t) dy + L Q'_{Ba}(t) \right]. \quad (8.6.9)$$

For the particular case under analysis, all the terminal voltages are constant. Therefore, (8.6.9) can be simplified to

$$I_G(t) = -W \frac{d}{dt} \int_0^L \frac{1}{n} Q'_l(y, t) dy. \quad (8.6.10)$$

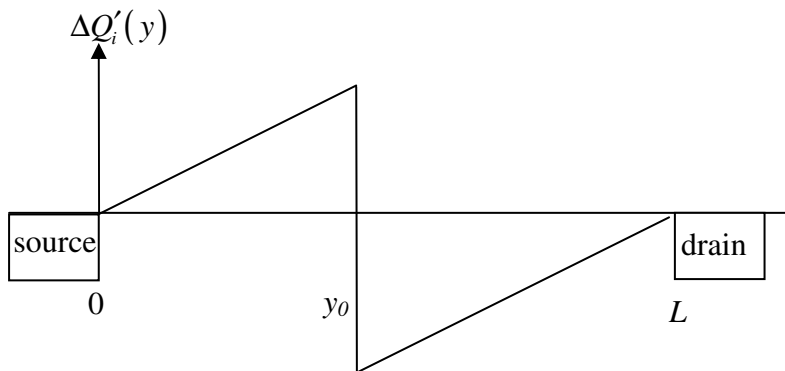


Fig. 8.6 Fluctuation in the inversion charge profile due to thermal voltage fluctuation at $y=y_0$.

The inversion charge density along the channel due to a jump $\Delta Q'_{i0}$ at $y=y_0$ is given by expressions (8.6.5) and (8.6.7); thus, the gate current in the frequency domain becomes

$$\begin{aligned} \Delta i_g &= -\frac{j\omega W}{n} \int_0^L \Delta Q'_i(y) dy \\ &= \frac{j\omega C'_{ox} \Delta i_d}{\mu} \left[\int_0^{y_0} \frac{y}{Q'_l(y)} dy + \int_{y_0}^L \frac{-(L-y)}{Q'_l(y)} dy \right] \quad (8.6.11) \\ &= \frac{j\omega C'_{ox} \Delta i_d}{\mu} \left[\int_0^L \frac{y}{Q'_l(y)} dy - \int_{y_0}^L \frac{L}{Q'_l(y)} dy \right]. \end{aligned}$$

In order to solve (8.6.11) in terms of the inversion charge density, let us recall (5.2.18), rewritten below as

$$y = \frac{\mu_n W}{2nC'_{ox} I_D} [Q'^2_{it}(0) - Q'^2_{it}(y)]. \quad (8.6.12)$$

The substitution of (8.6.12) into (8.6.11) and the integration along the channel yields

$$\Delta i_g = \frac{j\omega \Delta i_d WL}{3nI_D} \left[2Q'_{it}(0) \frac{1+\alpha+\alpha^2}{1+\alpha} - 3Q'_{it}(y) \right] \quad (8.6.13)$$

for a noise source located at a generic position y in the channel. Expression (8.6.13) gives the gate current Δi_g induced by the small perturbation Δi_d . We can proceed now to determine the total gate current by adding the contribution of all channel elements, *i.e.*

$$i_g = \sum_{channel} \Delta i_g. \quad (8.6.14)$$

In order to evaluate the mean square value of the gate current, we calculate the following summation along the channel

$$\overline{i_g^2} = \overline{\left(\sum_{channel} \Delta i_g \right)^2}. \quad (8.6.15)$$

Using (8.6.13) in (8.6.15) under the assumption that the elementary contributions of the channel current are uncorrelated we find that

$$\overline{i_g^2} = \left(\frac{\omega WL}{3nI_D} \right)^2 \sum_{channel} [H(y)]^2 \overline{\Delta i_d^2} \quad (8.6.16)$$

with

$$H(y) = 2Q'_{it}(0) \frac{1+\alpha+\alpha^2}{1+\alpha} - 3Q'_{it}(y). \quad (8.6.17)$$

Using (8.6.2), (8.6.1), and making $\Delta y \rightarrow 0$, the mean square value of the gate current results in

$$\frac{\overline{i_g^2}}{4kT\Delta f} = -\mu W \left(\frac{\omega W}{3nI_D} \right)^2 \int_0^L [H(y)]^2 Q'_t(y) dy. \quad (8.6.18)$$

Inserting (8.6.17) into (8.6.18) and solving the integral yields

$$\frac{\overline{i_g^2}}{4kT\Delta f} = \frac{16}{135} \frac{\omega^2 (C'_{ox} L)^3 W \phi_t}{\mu Q'^2_{lt}(0)} \left\{ \frac{Q'_{lt}(0)}{Q'_{lp}} \frac{1+5\alpha+10.5\alpha^2+5\alpha^3+\alpha^4}{(1+\alpha)^5} - \frac{15}{8} \frac{1+5\alpha+5\alpha^2+\alpha^3}{(1+\alpha)^5} \right\} \quad (8.6.19)$$

Expression (8.6.19) gives the mean square value of the induced gate noise as a function of both the inversion level at the source and the saturation index α for a long-channel device in quasi-static operation. Expression (8.6.19) is identical to the one given in reference [7].

In strong inversion saturation ($Q'_{lt}(0)/Q'_{lp} \gg 1$, $\alpha \ll 1$), expression (8.6.19) is approximated by

$$\frac{\overline{i_g^2}}{4kT\Delta f} \cong -\frac{16}{135} \frac{\omega^2 C'^2_{ox} W L^3}{\mu Q'_l(0)}, \quad (8.6.20)$$

which fully agrees with the result given by van der Ziel in [22]². One can also note that, in the linear region, for $V_{DS} \rightarrow 0$, or equivalently, $\alpha \rightarrow 1$, the induced gate noise is given by

$$\frac{\overline{i_g^2}}{4kT\Delta f} \cong -\frac{16}{135} \frac{\omega^2 C'^2_{ox} W L^3}{\mu Q'_l(0)} \frac{22.5}{32}, \quad (8.6.21)$$

which is slightly smaller than in saturation. In weak inversion, the mean square value of the gate current noise is

$$\frac{\overline{i_g^2}}{4kT\Delta f} = -\frac{16}{135} \frac{\omega^2 C'_{ox} W L^3}{\mu n^2 \phi_t^2} \frac{45}{128} Q'_l(0) = -\frac{1}{24} \frac{\omega^2 C'_{ox} W L^3}{\mu n^2 \phi_t^2} Q'_{IS} \quad (8.6.22)$$

for both linear and saturation regions.

Fig. 8.7 displays the normalized mean square value of the gate current noise in terms of the normalized inversion level at the source, having the

² Expression (5.66) in reference [22] is the same as (8.6.20), given that the drain conductance for $V_{DS}=0$ is $g_{d0} = -\mu Q'_l(0)W/L$.

ratio of the drain inversion charge to the source inversion charge as a parameter. In weak inversion the PSD of the gate current increases proportionally with the channel inversion charge owing to the increase in the elementary noise current with an increase in the charge. On the other hand, in strong inversion, the fluctuation in voltage drop (and in charge density) along the channel is less sensitive to the noise current owing to the lower resistance of the channel; consequently, the gate current is inversely proportional to the inversion charge of the channel.

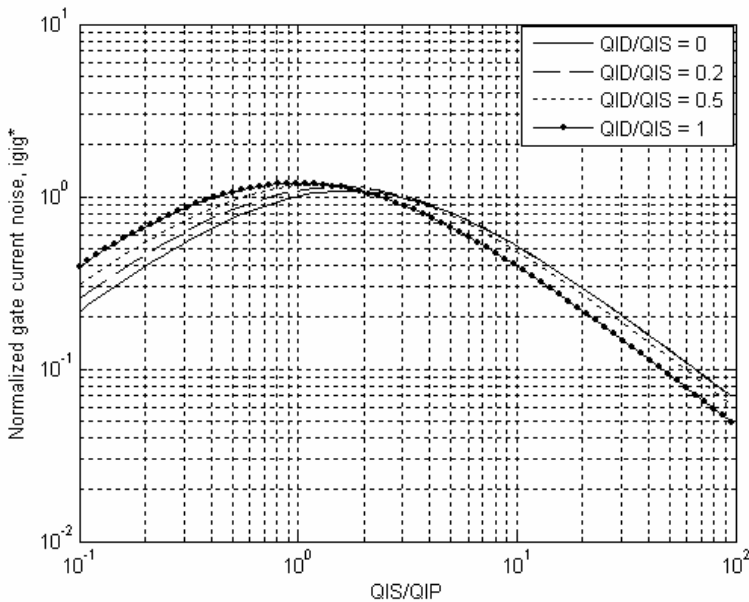


Fig. 8.7 Normalized power spectral density of the gate current noise. The values of the gate current noise are normalized to the gate current noise for $q'_{IS} = 1$, $q'_{ID} = 0$.

We now continue to calculate the correlation coefficient c between gate and drain noise

$$c = \frac{\overline{i_g i_d^*}}{\sqrt{\overline{i_g^2}} \sqrt{\overline{i_d^2}}}. \quad (8.6.23)$$

To obtain c , one needs to calculate the cross-spectral intensity $\overline{i_g i_d^*}$ from

$$\overline{i_g i_d^*} = \overline{\left(\sum_{channel} \Delta i_g \right) \left(\sum_{channel} \Delta i_d \right)^*}. \quad (8.6.24)$$

Since the noise sources along the channel are assumed to be uncorrelated, the substitution of (8.6.13) into (8.6.24) leads to

$$\overline{i_g i_d^*} = \sum_{channel} \frac{j\omega WL}{3nI_D} \left[2Q'_l(0) \frac{1+\alpha+\alpha^2}{1+\alpha} - 3Q'_l(y) \right] \overline{\Delta i_d \Delta i_d^*}. \quad (8.6.25)$$

Using (8.6.2), (8.6.1), and making $\Delta y \rightarrow 0$, and recalling that

$$I_D = \frac{\mu}{2nC'_{ox}} \frac{W}{L} Q'^2_l(0) (1-\alpha^2), \quad (8.6.26)$$

we find that

$$\begin{aligned} \frac{\overline{i_g i_d^*}}{4kT\Delta f} &= j \frac{4C'_{ox} WL \omega}{3Q'^4_l(0) (1-\alpha^2)^2} \\ &\int_{Q'_l(0)}^{Q'_l(L)} Q'_l(y) \left[2Q'_l(0) \frac{1+\alpha+\alpha^2}{1+\alpha} - 3Q'_l(y) \right] Q'_l(y) dQ'_l, \end{aligned} \quad (8.6.27)$$

which integrated along the channel gives

$$\frac{\overline{i_g i_d^*}}{4kT\Delta f} = \frac{1}{9} j\omega (C'_{ox} WL) \left[\frac{(1-\alpha)(1+4\alpha+\alpha^2)}{(1+\alpha)^3} \right]. \quad (8.6.28)$$

Now, using (7.2.7) (with Q_l given in terms of α as in (5.2.22)), (8.6.19), and (8.6.28) we find the correlation coefficient between gate and drain thermal noise, which is represented in Fig. 8.8.

$$c = \left[\frac{j\sqrt{\frac{5}{32}}(1-\alpha)(1+4\alpha+\alpha^2)\frac{Q'_h(0)}{Q'_{IP}}}{\sqrt{\frac{Q'_h(0)}{Q'_{IP}}[1+5\alpha+10.5\alpha^2+5\alpha^3+\alpha^4]} - \frac{15}{8}[1+5\alpha+5\alpha^2+\alpha^3]}} \right] \times \frac{1}{\sqrt{\frac{Q'_h(0)}{Q'_{IP}}(1+\alpha+\alpha^2) - \frac{3}{2}(1+\alpha)}}} \quad (8.6.29)$$

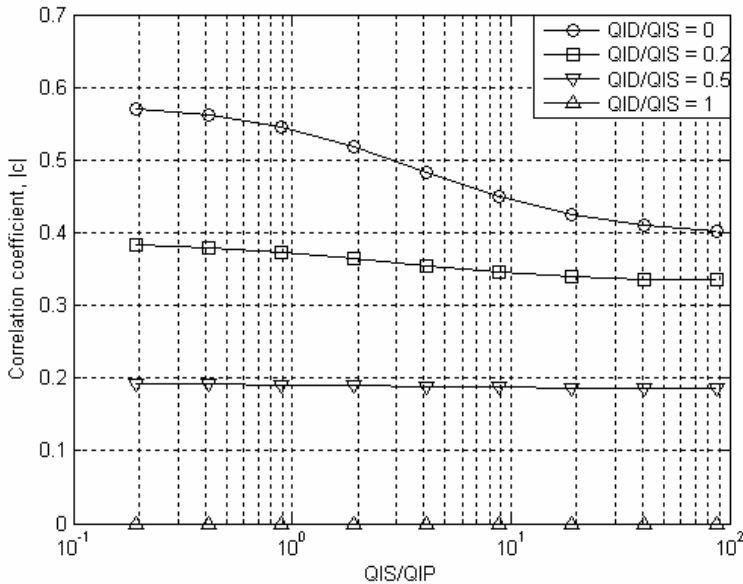


Fig. 8.8 Correlation coefficient of gate and drain noise vs. normalized inversion charge density at source from the linear region, $q'_{ID}/q'_{IS}=1$, to the saturation region, $q'_{ID}/q'_{IS}=0$.

In strong inversion saturation ($Q'_h(0)/Q'_{IP} \gg 1$, $\alpha \ll 1$), the correlation coefficient is equal to $j\sqrt{5/32} \cong j0.395$ as given in [22]. In

weak inversion and saturation, $c \rightarrow j\sqrt{3}/3 \cong j0.58$. Our result for the correlation coefficient is in full agreement with that given in [7] and [8]. At first sight, one could infer that lcl would be very close to 1. However, as commented by van der Ziel in [23], the contributions of the different segments of the channel to the cross-correlation between gate and drain noise partly cancel each other.

Figure 8.8 shows the magnitude of the correlation coefficient of the drain to gate noise. For very low V_{DS} , $lcl \rightarrow 0$. In saturation, $lcl \rightarrow 0.395$ in strong inversion and $lcl \rightarrow 0.58$ in weak inversion.

The model we have developed for the induced gate noise assumes the gate current to be significantly smaller than the channel current. Using the expressions for the channel thermal noise and the gate induced noise, it can be demonstrated that, in strong inversion, the assumption of small gate current is valid as long as the operating frequency is smaller than the intrinsic MOSFET transition frequency.

Bibliography

- [1] Yannis Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed., McGraw-Hill, Boston, 1999
- [2] William Liu, *MOSFET Models for SPICE Simulation including BSIM3v3 and BSIM4*, John Wiley & Sons, New York, 2001.
- [3] Behnam Aghdaie and Bing Sheu, "Overcoming limitations of lumped MOS models," *IEEE Circuits Devices Mag.*, vol. 16, no. 2, pp. 19-26, Mar. 2000.
- [4] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, Physical background of MOS model 11, Nat. Lab. Unclassified Report 2003/00239. [Online]. Available: http://www.semiconductors.philips.com/Philips_Models/
- [5] Mehran Bagheri and Yannis Tsividis, "A small signal dc-to-high-frequency nonquasistatic model for the four-terminal MOSFET valid in all regions of operation," *IEEE Trans. Electron Devices*, vol. 32, no. 11, pp. 2383-2391, Nov. 1985.
- [6] Mehran Bagheri, *Solutions manual to accompany Tsividis Operation and Modeling of the MOS Transistor*, McGraw-Hill, New York, 1987.
- [7] A.-S. Porret and C. C. Enz, "Non-quasi-static (NQS) thermal noise modelling of the MOS transistor," *IEE Proc.-Circuits Devices Syst.*, vol. 151, no. 2, pp. 155-166, Apr. 2004.
- [8] Alain-Serge Porret, *Design of a Low-Power and Low-Voltage UHF Transceiver Integrated in a CMOS Process*, PhD Thesis, EPFL, Lausanne, Switzerland, 2002.
- [9] Narain Arora, *MOSFET Models for VLSI Circuit Simulation Theory and Practice*, Springer-Verlag, Wien, 1993.

- [10] Paolo Mancini, Claudio Turchetti, and Guido Masetti, "A non-quasi-static analysis of the transient behavior of the long channel MOST valid in all regions of operation," *IEEE Trans. Electron Devices*, vol. 34, no. 2, pp. 325-334, Feb. 1987.
- [11] Soo-Young Oh, Donald E. Ward, and Robert W. Dutton, "Transient analysis of MOS transistors," *IEEE Trans. Electron Devices*, vol. 27, no. 8, pp. 1571-1578, Aug. 1980.
- [12] D. H. Treleaven and F. N. Trofimenkoff, "MOS FET equivalent circuit at pinch-off," *Proc. IEEE*, vol. 54, no. 9, pp. 1223-1224, Sep. 1966.
- [13] J. M. Sallese and A.-S. Porret, "A novel approach to charge-based non-quasi-static model of the MOS transistor valid in all modes of operation," *Solid-State Electron.*, vol. 44, no. 6, pp. 887-894, June 2000.
- [14] J. A. van Nielen, "A simple and accurate approximation to the high-frequency characteristics of insulated-gate field-effect transistors," *Solid-State Electron.*, vol. 12, no. 10, pp. 826-829, Oct. 1969.
- [15] Ana Isabela Araújo Cunha, *Um Modelo do Transistor MOS para Projeto de Circuitos Integrados*, PhD Thesis, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 1996 (in Portuguese). [Online]. Available: <http://eel.ufsc.br/~lci/publications.html>
- [16] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510-1519, Oct. 1998.
- [17] Carlos Galup-Montoro, Márcio Cherem Schneider, and Ana Isabela Araújo Cunha, "A current-based MOSFET model for integrated circuit design," Chapter 2 in *Low-Voltage/Low-Power Integrated Circuits and Systems*, pp. 7-55, edited by Edgar Sánchez-Sinencio and Andreas G. Andreou, IEEE Press, Piscataway, NJ, 1999.
- [18] A. J. Scholten, L. F. Tiemeijer, P. W. H. de Vreede, and D. B. M. Klaassen, "A large signal non-quasi-static MOS model for RF circuit simulation," *IEDM Tech. Dig.*, 1999, pp. 163-166.
- [19] Andries J. Scholten, Luuk F. Tiemeijer, Ronald van Langevelde, Ramon J. Havens, Adrie T. A. Zegers-van Duijnhoven, and Vincent C. Venezia, "Noise modeling for RF CMOS circuit simulation," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618-632, Mar. 2003.
- [20] A. van der Ziel, "Gate noise in field effect transistors at moderately high frequencies," *Proc. IEEE*, vol. 51, no. 3, pp. 461-467, Mar. 1963.
- [21] Masakazu Shoji, "Analysis of high-frequency thermal noise of enhancement mode MOS field-effect transistors," *IEEE Trans. Electron Devices*, vol. 13, no. 6, pp. 520-524, June 1966.
- [22] Aldert van der Ziel, *Noise in Solid State Devices and Circuits*, Wiley-Interscience, New York, 1986.
- [23] Aldert van der Ziel, "Noise in solid-state devices and lasers," *Proc. IEEE*, vol. 58, no. 8, pp. 1178-1206, Aug. 1970.

Problems

8.1. Derive expression (8.3.2). Plot the inversion charge density, normalized to the inversion charge density at the source, along the channel of a saturated transistor, for $q'_{IS} = 0.1, 1$, and 10 .

8.2. Verify that the transimpedance $-z_{gs} = -1/y_{gs}$ can be represented as the series association of capacitor C_{gs} and a resistance R_{gs} for $\omega\tau_2 \ll 1$. (Hint: Use the approximation $(1 + j\omega\tau_2)^{-1} \cong 1 - j\omega\tau_2$.) Demonstrate that for strong inversion saturation, $R_{gs} \cong 1/5g_{mg}$.

8.3. Let us define $\tilde{\omega}_T = g_{mg}/(C'_{ox}WL/n) \cong \omega_T$, which is approximately equal to the intrinsic cutoff frequency. Show that, in strong inversion saturation, the ratio of the PSD of the induced gate noise to the PSD of the channel noise is given by

$$\frac{S_{I_g}}{S_{I_d}} \cong \frac{2}{45} \left(\frac{\omega}{\tilde{\omega}_T} \right)^2. \quad (\text{P8.3.1})$$

Chapter 9

Gate and Bulk Currents

For oxides below 4nm, high current leakages through the oxide can occur due to the quantum mechanical tunneling of electrons. The gate leakage current can not only negatively affect the device performance [28] but can also significantly increase the standby power consumption of a chip. For these reasons a compact model for the gate current is mandatory for advanced technologies. Although the compact modeling of the gate tunneling currents is a relatively new subject, the characterization and modeling of gate tunneling currents is not. Beginning with the pioneering work of Lenzlinger and Snow [1], numerous researchers have analyzed electron tunneling current in MOS structures, but most of the literature has focused on thick oxides and high applied voltages. The tunneling models for thick oxides are not adequate for thin oxides because of the differences in the potential barrier shapes and transport mechanisms.

In this chapter we will first briefly review tunneling through a potential barrier. A compact model for tunneling in MOS structures is then developed and a brief comparison of the main tunneling models is carried out. The last topic of the gate current section is the analysis of the different components of the gate current. The other subject of the chapter is the bulk current. The two main mechanisms at the origin of bulk current, namely gate-induced drain leakage (GIDL), and impact ionization current, are presented.

9.1 Gate tunneling current

Before developing the tunneling current model for the MOS structure we will briefly review the basics of quantum mechanical tunneling.

9.1.1 Tunneling through a potential barrier [2]

Tunneling is the quantum-mechanical process of the passage of a particle through a classically forbidden region of space. The analysis of tunneling is based on the solution of the Schrödinger equation, which describes the wave function, associated with the probability of finding an electron in a given region of space.

Consider the simple case of an electron tunneling through a rectangular potential barrier. The wave nature of the electron implies a non zero probability of the electron penetrating the barrier as shown in Fig. 9.1, even if the kinetic energy E of the electron is lower than the potential barrier height V_o . The electron wave function penetrates the barrier region, and if the barrier thickness a is thin enough compared with the de Broglie wavelength ($\lambda = h/p$, h is the Planck constant and p is the electron momentum), a non negligible probability of finding the electron on the other side of the barrier results. The analysis of a particle tunneling through a rectangular barrier is one of the basic textbook examples for the solution of the Schrödinger equation. The transmission coefficient T for a particle tunneling through a rectangular barrier of width a is given by

$$T \cong T_0 \exp \left\{ -\frac{p_{mis} a}{\hbar/2} \right\} \quad (9.1.1)$$

where p_{mis} is the momentum associated with the energy the particle is lacking ('missing') in order to penetrate the barrier region, $\hbar = h/2\pi$ is the reduced Planck constant and $T_0 = 16E(V_0 - E)/V_0^2$. p_{mis} is given by

$$p_{mis} = \sqrt{2m(V_0 - E)} \quad (9.1.2)$$

where m is the mass of the particle. Comparing the exponent in (9.1.1) with the uncertainty principle, (9.1.3), reveals that the tunneling process is closely related to the uncertainty principle [2].

$$\frac{\Delta p \Delta x}{\hbar/2} \geq 1. \quad (9.1.3)$$

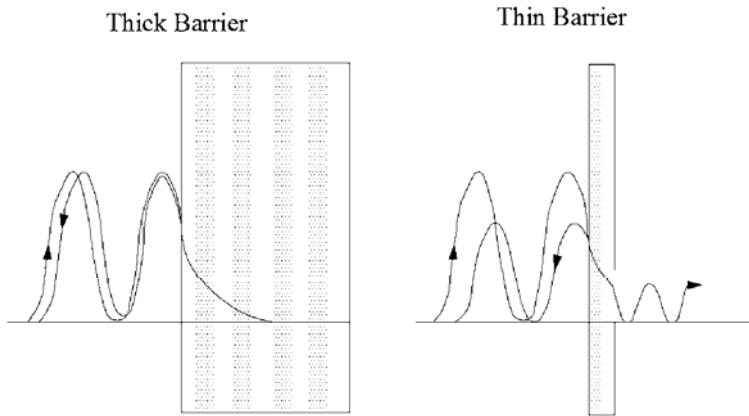


Fig. 9.1 Representation of the incident, reflected and transmitted wave functions of an electron tunneling through thick and thin barriers.

An approximate expression for the transmission coefficient for a potential barrier with an arbitrary shape can be regarded as a generalization of (9.1.1) and is

$$T = T_0 \exp\left\{-2 \int_c^d K(x) dx\right\}, \quad (9.1.4)$$

where

$$K(x) = \frac{p_{mis}(x)}{\hbar} = \frac{\sqrt{2m(V(x) - E)}}{\hbar} \quad (9.1.5)$$

and c and d are the limits of the forbidden region ($V(c)=V(d)=E$). The exponential term gives the attenuation of the wave function in the forbidden region. If T_0 in (9.1.4) is made equal to one, then (9.1.4) is called the Wentzel-Kramers-Brillouin (WKB) approximation for the tunneling probability. T may also be approximated by substituting $V(x)$ with the maximum value of the potential in the barrier [2].

9.1.2 Compact model for tunneling in MOS structures

As shown in Fig. 9.2 [3], there are several paths for electrons and holes to tunnel through the oxide of an MOS structure.

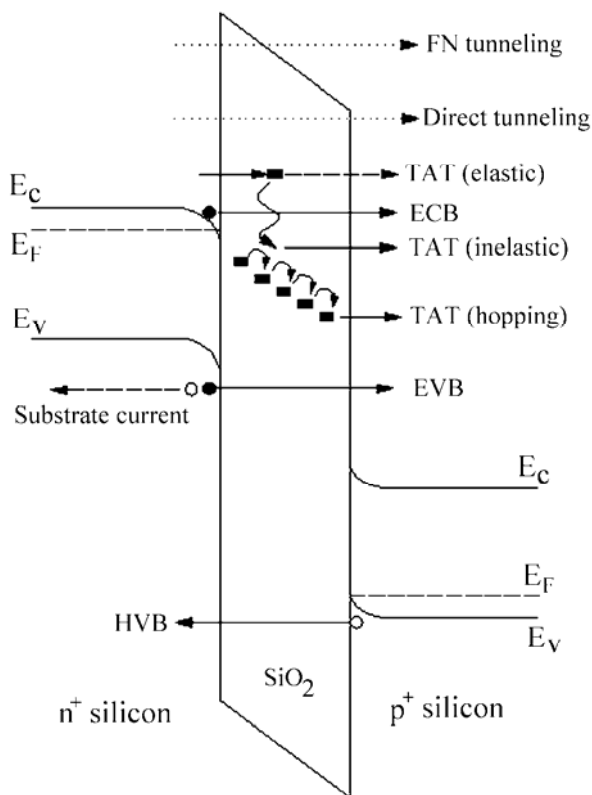


Fig. 9.2 Tunneling mechanisms, according to either bias-dependent classification (Fowler-Nordheim and direct tunneling) or systematic classification (ECB – conduction band electrons, HVB – valence band holes, EVB – interband tunneling, TAT – trap-assisted tunneling). (After [3].)

For electrons, the two main processes are tunneling through a triangular barrier, called Fowler-Nordheim (FN) tunneling, and tunneling through a trapezoidal barrier, called direct tunneling. For oxide potential drops V_{ox} greater than the Si-SiO₂ conduction band discontinuity $\chi_B \cong 3.15$ V, as is usual for thick oxides, FN tunneling is predominant. On the other hand, for thin oxides ($t_{ox} \leq 3 \text{ nm}$) direct tunneling prevails and the gate current can be significant, even for V_{ox} substantially less than χ_B .

A compact gate tunneling current model requires the calculation of the charge sheet density Q'_I available for tunneling and the average carrier tunneling lifetime τ . The tunneling current density can be simply given [4] by

$$J_G = Q'_I / \tau . \quad (9.1.6)$$

In the case of substrate injection, Q'_I corresponds to the accumulation or the inversion charge. Both Q'_I and τ depend on bias. The value of the inversion charge density is calculated in the same way as for the conventional MOS model, considering (quasi)equilibrium. In other terms, the number of carriers lost by tunneling is considered not to alter significantly the inversion (accumulation) carrier densities. The tunneling lifetime is related to the transparency coefficient of the barrier (or transmission probability) T and to the frequency f of the impact against the barrier by [4]

$$1/\tau = fT . \quad (9.1.7)$$

The impact frequency can be regarded as the number of collisions of a particle against the barrier per unit time.

Alternatively we can rewrite (9.1.6) as

$$J_G = Q'_I fT . \quad (9.1.8)$$

After introducing (9.1.4) to calculate the transmission probability, the only unknown term for the calculation of the tunneling current is the impact frequency.

9.1.2.1 Impact frequency [5], [6]

Let us consider a carrier confined in a potential well of depth l . To calculate the impact frequency in a simple way, let us assume that the particle is confined to moving in the x-direction only; therefore, the momentum of the particle has a component in the x-direction only. If $v(x)$ is its velocity it will take the time

$$t = \int_0^l v^{-1} dx \quad (9.1.9)$$

to cross the well. Consequently, the impact frequency is given by

$$f^{-1} = 2 \int_0^l v^{-1} dx. \quad (9.1.10)$$

Here, we recall that the kinetic energy E is related to the momentum and to the wave number through

$$\frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} = E. \quad (9.1.11)$$

The simplest quantum mechanical model for electrons at the interface is an infinitely deep unidimensional well of depth l . The allowed energy states for the confined electrons are quantized as

$$E_n = \frac{\hbar^2}{2m} \left(\frac{n\pi}{l} \right)^2. \quad (9.1.12)$$

Considering the electron in the lowest energy level ($n=1$), it follows that the “average” velocity in the x-direction is

$$v = \sqrt{\frac{2E_1}{m}} \quad (9.1.13)$$

where E_1 is the energy of the lowest sub-band in the x-direction. Substituting (9.1.13) into (9.1.10) yields

$$f = \sqrt{\frac{E_1}{2ml^2}}. \quad (9.1.14)$$

Finally, writing l in terms of E_1 using (9.1.12) it follows that

$$f = \frac{2E_1}{h}. \quad (9.1.15)$$

The very simple approximations that have been used here to derive (9.1.15) should be modified for a more realistic confinement model of electrons.

9.1.2.2 Equations for the tunneling current

Since each of the terms Q'_I , f , and T in the expression of the gate current, (9.1.8), can be modeled in different ways, there are a great variety of tunneling models available in the literature.

Until recently, because of the relatively thick oxide employed in the MOS structures, researchers were mostly interested in tunneling in strong inversion and accumulation, and consequently the charge available for tunneling was modeled simply as

$$-Q'_I \cong \epsilon_{ox} F_{ox} = \epsilon_{ox} V_{ox} / t_{ox} . \quad (9.1.16)$$

Several of the classical MOS tunneling current formulas have an F_{ox}^2 pre-exponential term, due to the tunneling charge and the pre-exponential factor of the impact frequency both being proportional to F_{ox} . The most important of these formulas, the FN tunneling formula (9.1.17), first derived to explain the field emission of electrons from a metal electrode under vacuum, gives an excellent fit for the gate current in thick oxide MOS structures as shown in Fig. 9.3.

$$J_g = CF_{ox}^2 \exp\{-\beta/F_{ox}\} . \quad (9.1.17)$$

C and β are physical parameters. As previously mentioned, to derive (9.1.17) it is considered that the electrons tunnel through a triangular barrier.

Expression (9.1.18) [7] was derived considering the trapezoidal barrier shape for thin oxides instead of the triangular barrier shape.

$$J_g = DF_{ox}^2 \exp\left\{-\frac{B}{F_{ox}} \left[1 - \left(1 - \frac{V_{ox}}{\chi_B}\right)^{3/2}\right]\right\} . \quad (9.1.18)$$

χ_B is the Si-SiO₂ conduction band discontinuity. The BSIM4 model [8] uses a slightly modified version of formula (9.1.18).

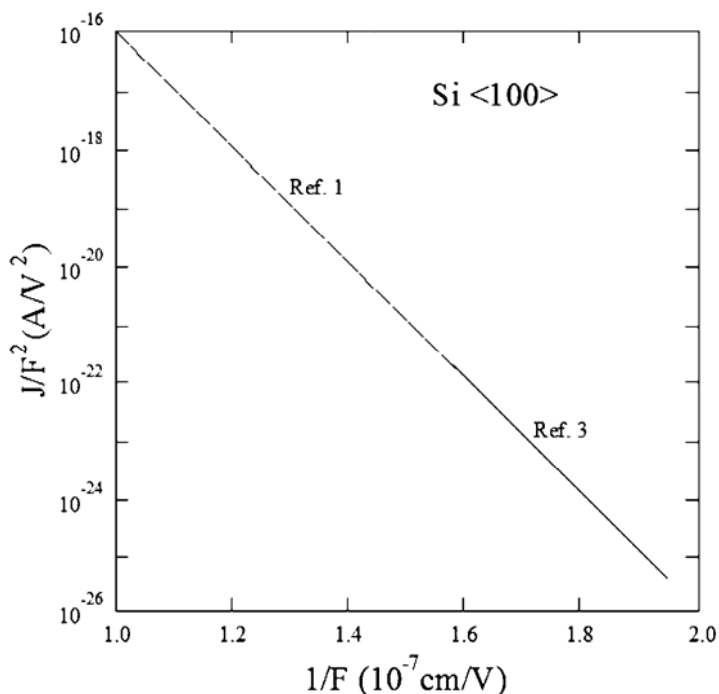


Fig. 9.3 Fowler-Nordheim plots of tunneling currents from Si <100> into SiO_2 . The solid line, labeled Ref. 3, corresponds to samples with 100nm of SiO_2 on n-type Si. The dashed line, labeled Ref. 1, corresponds to samples with 7nm of SiO_2 on n-type Si. (After [10].)

As noted previously, the two formulas above are valid only for strong inversion, because in the two cases the inversion channel charge is considered proportional to F_{ox} . Recent studies consider a general model of the carrier charge valid in the different operating regions. Regarding the terms T and f , a lot of variations are possible [5], [6], [9].

To use expression (9.1.15) for the impact frequency we need to calculate the first energy level for the electron at the interface. For a triangular potential well the energy levels [10], [11] are

$$E_n = \frac{Z_n}{2^{1/3}} \left(\frac{q\hbar}{m_s^{1/2}} F_s \right)^{2/3}, \quad (9.1.19)$$

where the Z_n 's are the zeros of the Airy function, m_s is the electron mass in the semiconductor, and F_s is the electric field in the inversion region. Using (9.1.15), (9.1.19), and the simplest model for the transparency coefficient with a constant pre-exponential term (the WKB approximation) yields

$$fT \propto V_{ox}^{2/3} \exp \left\{ -2 \int_c^d K(x) dx \right\}. \quad (9.1.20)$$

A similar and simpler result in [12] gives $f.T \propto E_1^{3/2}$ and, consequently, V_{ox} appears as the pre-exponential term.

Expression (9.1.21) adopted in MM11 [13] has the V_{ox} pre-exponential term of Weinberg's formula and the value of the tunneling integral $-2 \int_c^d K(x) dx$ adopted in BSIM.

$$f.T = AV_{ox} \exp \left\{ -\frac{B}{V_{ox}} \left[1 - \left(1 - \frac{V_{ox}}{\chi_B} \right)^{3/2} \right] \right\} \quad (9.1.21)$$

where A and B are physical parameters given by

$$A = \frac{q}{m_s} \frac{\epsilon_{ox}}{\epsilon_s} \sqrt{\frac{2m_{ox}}{q\chi_B}} \frac{1}{t_{ox}} \quad (9.1.22)$$

$$B = \frac{4}{3} \frac{\sqrt{2qm_{ox}}}{\hbar} t_{ox} \chi_B^{3/2} \quad (9.1.23)$$

and m_{ox} is the electron effective mass in the oxide.

9.1.2.3 Comparison between models [14]

Very different approaches to tunneling current in MOS structures have been developed, from the fundamental approach of Bardeen [15] to the transparency methods summarized in the last sections. It has been recently stated [14] that all these methods, despite their very different formalisms, are based on similar physical assumptions and give very close results. Fig. 9.4 gives the tunneling lifetime following the main approaches. The discrepancies between the different approximations can

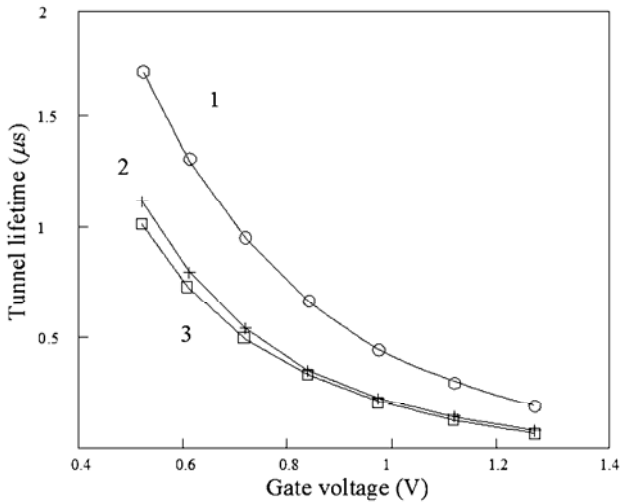


Fig. 9.4 Tunnel lifetime (of the ground state) of a 1.5nm oxide vs. gate voltage. (1) model using Weinberg formula for impact frequency and WKB transparency, (2) numerical Bardeen's approach, and (3) model based on improved WKB transparency expression. (After [14].)

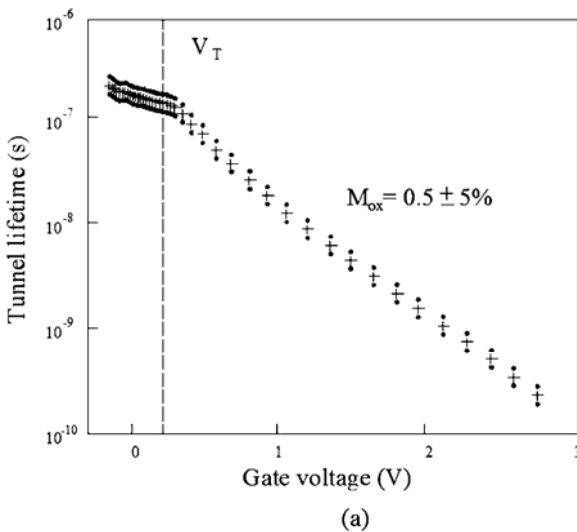


Fig. 9.5 Impact of the main parameters (normalized electron effective mass M_{ox} in the oxide and oxide thickness) on the tunnel lifetime. (After [14]).

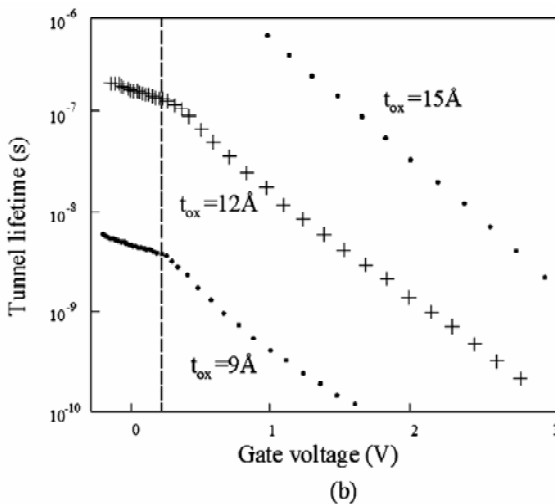


Fig. 9.5 (Continued)

be considered negligible when compared to the high sensitivity of the tunneling current expression to changes in physical parameters such as the effective mass or oxide thickness as demonstrated in Fig. 9.5.

9.1.2.4 Partition of the gate current into the source and drain components

For an nMOS transistor operating in inversion, the intrinsic gate current is due to electrons tunneling from the inversion layer. The gate current density satisfies the continuity equation, given by

$$WJ_G(y) = \frac{\partial I}{\partial y} \quad (9.1.24)$$

where I is the channel current. Equation (9.1.24) is similar to the continuity equation of (5.1.3). In fact, if we substitute J_G in (9.1.24) with $\partial Q'_1 / \partial t$ we obtain Eq. (5.1.3). The Ward-Dutton method, presented in Section 5.1.2 to determine the source and drain components of the charging currents, is based on the current continuity equation (5.1.3). Consequently, we can also apply the Ward-Dutton method to the splitting of the gate current between source and drain. The splitting of the charging currents, calculated as derivatives of the source and drain

charges, results from the splitting of the total inversion charge between source and drain given by Eqs. (5.1.12) and (5.1.14) repeated below

$$Q_S = W \int_0^L \left(1 - \frac{y}{L}\right) Q'_I dy \quad (9.1.25)$$

$$Q_D = W \int_0^L \frac{y}{L} Q'_I dy. \quad (9.1.26)$$

Changing the notation as indicated below

$$\frac{\partial Q'_I}{\partial t} \rightarrow J_G \quad \frac{dQ_{S(D)}}{dt} \rightarrow I_{GS(D)}, \quad (9.1.27)$$

equations (9.1.25) and (9.1.26) are rewritten as

$$I_{GS} = W \int_0^L \left(1 - \frac{y}{L}\right) J_G dy \quad (9.1.28)$$

$$I_{GD} = W \int_0^L \frac{y}{L} J_G dy. \quad (9.1.29)$$

Equations (9.1.28) and (9.1.29) were first presented in [16] and [17]. The total intrinsic gate current is given by

$$I_G = I_{GS} + I_{GD} = W \int_0^L J_G dy. \quad (9.1.30)$$

In accumulation, the intrinsic gate current consists of electrons tunneling from gate to bulk. Because no channel current is flowing, the calculation of the gate-to-bulk current is straightforward, given simply by

$$I_{GB} = W \int_0^L J_{GB} dy. \quad (9.1.31)$$

Overlap currents can be included considering these regions as specific MOS structures with their specific flat-band voltage and body factor, see for example [13]. Including all the above components, the gate current model of MM11 gives an accurate description of I_G , as shown in Fig. 9.6.

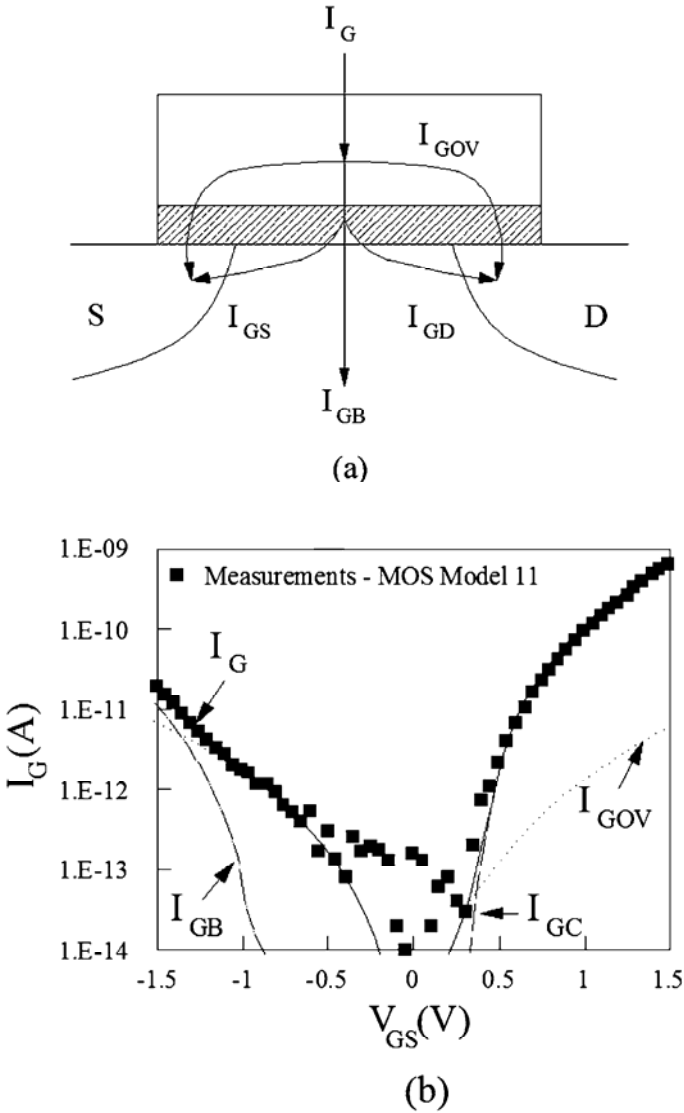


Fig. 9.6 (a) The different gate current components in a MOSFET. (b) Measured and modeled gate current as a function of gate bias V_{GS} for $V_{DS}=V_{SB}=0$ V. (After [13].)

9.2 Bulk current

In addition to the bulk current occasionally resulting from tunneling through the gate oxide, other mechanisms can create bulk currents. Clearly, the drain-bulk and source-bulk diodes contribute with their reverse currents, but here we will restrict our analysis to the intrinsic transistor. Two mechanisms, namely band-to-band tunneling (BBT) under the drain-to-gate overlap region and impact ionization, can contribute to the bulk current.

9.2.1 Gate-induced drain leakage

A breakdown phenomenon at drain voltages much lower than the junction breakdown is detectable in very-thin-gate-oxide MOSFETs [18], [19], [20]. This breakdown is caused by the gate-induced high electric field in the gate-to-drain overlap region. This phenomenon was initially called subbreakdown but it is now recognized as gate-induced drain leakage (GIDL). This leakage current can be the dominant drain leakage current at zero gate bias in submicron devices.

For a negative gate-drain bias V_{GD} , a depletion region is formed under the gate-to-drain overlap region and the high electrical field in that region can make electrons of the valence band tunnel into the conduction band. As a consequence, electron-hole pairs are generated, the electrons being collected by the drain and the holes by the substrate, as shown in Fig. 9.7. Owing to its origin, this leakage current is called gate-induced drain leakage.

The GIDL current can be modeled by the formula given in [13], which is similar to the FN tunneling expression

$$I_{GIDL} \propto F_{ov}^2 \exp\left(-\frac{F_{GIDL}}{F_{ov}}\right) \quad (9.2.1)$$

where F_{GIDL} is an empirical parameter and F_{ov} is the maximum electric field in the gate oxide that overlaps the drain region.

Figure 9.8 shows typical transfer characteristics of an 88 Å gate oxide MOS transistor. The subthreshold current decreases exponentially

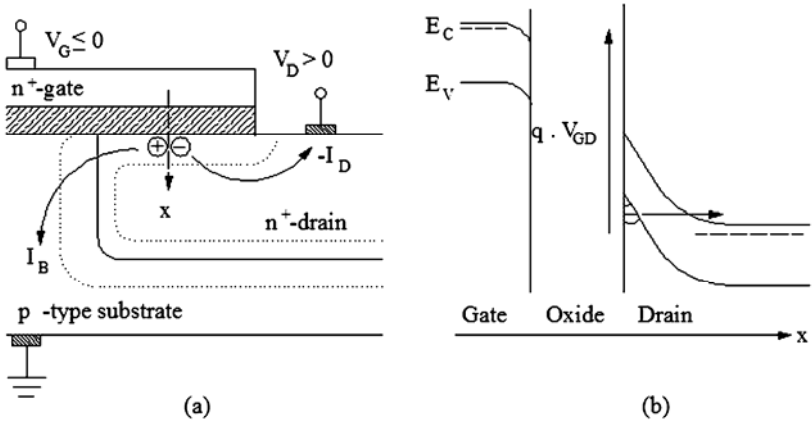


Fig. 9.7 (a) Cross-section of the gate-drain overlap region and (b) corresponding energy-band diagram along the x-direction. (After [13].)

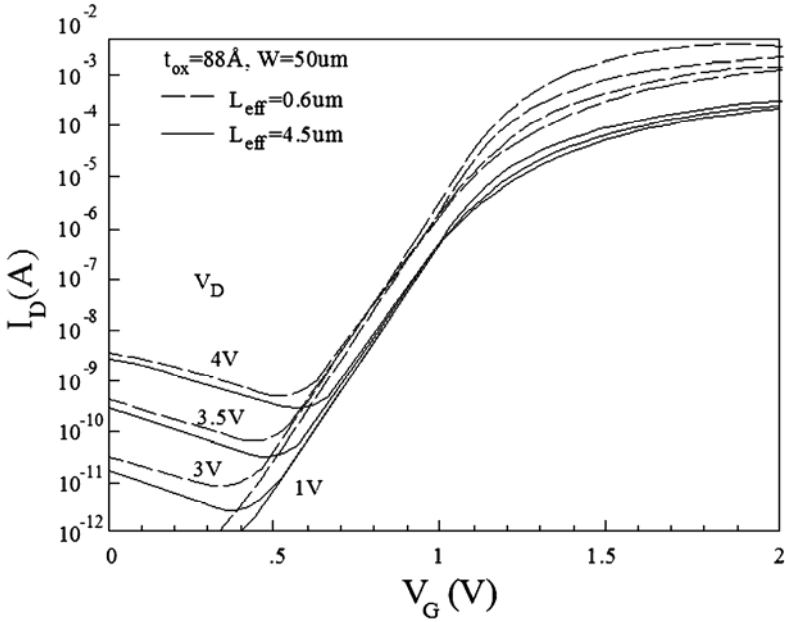


Fig. 9.8 Drain current characteristics for a short and a long channel nMOSFET. (After [18].)

with decreasing gate voltage and were it not for GIDL the current would decrease exponentially with gate voltage down to the level of the drain-substrate leakage current. The strong dependence of GIDL on the drain voltage, as predicted by (9.2.1), is also noticeable, which shows that GIDL is highly dependent on the electric field and, consequently, on the gate-to-drain voltage.

9.2.2 Impact ionization current

Impact ionization is a well studied process because it is one of the most important mechanisms in junction breakdown. As shown in Fig. 9.9, as the electrons move in the high field region near the drain, electron-hole pairs can be generated by impact ionization if the applied bias is strong enough. The electrons generated by impact increase the drain current and the holes are collected by the substrate, resulting in a substrate current.

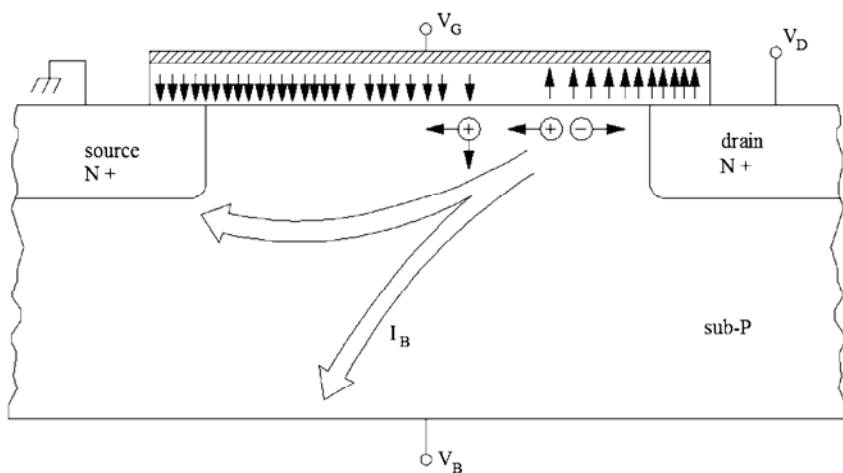


Fig. 9.9 Cross-section of an n-channel MOSFET. (After [23].)

9.2.2.1 Weak avalanche

In the case under consideration, far from the junction breakdown voltage, only low-level avalanche multiplication occurs; consequently, the calculation of the ionization integral [21] is greatly simplified.

In order to derive a model for the avalanche phenomenon in the MOSFET, we will make some simplifying assumptions. First, let us assume that the net motion of both electrons and holes is parallel to the semiconductor-oxide interface, as shown in Fig. 9.10, which displays a small channel element and the impact ionization phenomenon. Assuming dc-current only, the continuity equation is written [22] as

$$-\frac{dI_p}{dy} = \frac{dI_n}{dy} = \alpha_n I_n + \alpha_p I_p. \quad (9.2.2)$$

The first equality in (9.2.2) means that the sum of the hole and electron currents is constant along the region under consideration, as dictated by charge conservation. The two terms in the right-hand side are associated with the generation by impact of electron-hole pairs by either electrons (first term) or holes (second term), as illustrated in Fig. 9.10. The ionization rates α_n and α_p are the number of electron-hole pairs produced by an electron and a hole, respectively, per centimeter traveled in the direction of the electric field. The general solution for equation (9.2.2) can be found in [22].

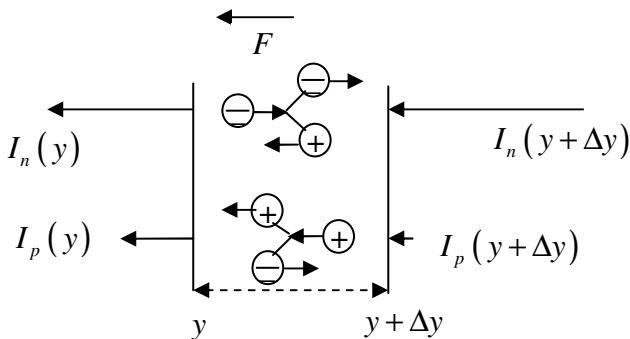


Fig. 9.10 Generation of electron-hole pairs by impact and electron and hole components of the current.

In the following, we will solve differential equation (9.2.2) for the particular case of a MOSFET. To that end, we split the channel into two parts, one closer to the drain, where high fields are responsible for impact

generation of electron-hole pairs and a region closer to the source, where the ionization rates and the recombination rates are negligible. We will assume that the potential barrier at the junctions prevents secondary holes from reaching either the source or drain diffusions [24]. Therefore, the secondary holes produced in the high-field region are collected in the substrate, as illustrated in Fig. 9.9. We now apply the concept in expression (9.2.2) to the saturation region of the MOSFET, closer to the drain. For the n-channel device, pure electron current initiates the multiplication, which implies that the hole current at the drain end of the channel is zero. In the analysis here, we shall be concerned with the particular case of low-level multiplication, where generation of electron-hole pairs by secondary holes is negligible compared to generation by primary electrons [24]. Therefore, for low-level multiplication we can apply expression (9.2.2) to Fig. 9.10, which results in

$$\frac{dI_p}{dy} = -\frac{dI_n}{dy} \cong -\alpha_n I_n \cong -\alpha_n I_n(L). \quad (9.2.3)$$

since $I_n \gg I_p$ and, for silicon, $\alpha_n > \alpha_p$. Note also that $I_n(L) = I_D$. The boundary condition for solving (9.2.3) is $I_p(L) = 0$ or $I_n(L) = I_n(y) + I_p(y)$. The integration of (9.2.3) along the high-field region $[L - \Delta L, L]$ yields

$$1 - \frac{1}{M_n} = \int_{L-\Delta L}^L \alpha_n dy, \quad (9.2.4)$$

where $M_n = I_n(L)/I_n(L - \Delta L)$ is the multiplication factor of electrons. The integral in (9.2.4) represents the fraction of the drain current which is injected into the bulk, that is

$$I_B = \frac{M_n - 1}{M_n} I_D. \quad (9.2.5)$$

The ionization rate of electrons α_n is a strong function of the electric field F . One of the most commonly used formula for it [25], [26] is

$$\alpha_n = A_i \exp\left(-\frac{F_i}{F}\right), \quad (9.2.6)$$

where A_i and F_i are called the impact ionization constants and F is the electric field. Values of $A_i=2.45 \times 10^{-6} \text{ cm}^{-1}$ and $F_i=1.92 \times 10^6 \text{ V/cm}$ have been experimentally determined for electrons at the silicon surface [25].

Due to the exponential dependence of α_n on electric fields, impact-ionization dominates at the region where the electric field is at its maximum. The calculation of the maximum field in the MOSFET requires a two-dimensional model. Using a compact quasi-two dimensional model, reference [26] demonstrates that (9.2.5) can be approximated by

$$I_B \cong I_D \frac{A_i (V_{DS} - V_{DSsat})}{F_i} \exp\left(-\frac{lF_i}{V_{DS} - V_{DSsat}}\right) \quad (9.2.7)$$

where l is an effective ionization length, given by $l^2 = (\epsilon_s / \epsilon_{ox}) t_{ox} X_j$ and X_j is the junction depth.

Figure 9.11 shows the two components used to evaluate the substrate current. The first component is the channel current which is an increasing function of the gate voltage. The second component is the multiplication factor which decreases with the gate voltage for a drain voltage held constant. In effect, as V_{GS} increases, so does V_{DSsat} , and this in turn reduces the maximum electric field. The product of the two components produces a substrate current with a peak as shown in Fig. 9.11. Experimental and simulation results for the substrate impact ionization current are given in Fig. 9.12.

For advanced CMOS technologies, because the supply voltage is of the same order as, or even lower than the band-gap of silicon, the carriers are no longer able to attain enough energy to create electron-hole pairs. Consequently, impact ionization is no longer important for advanced silicon technologies.

Finally, the bulk current for an nMOS transistor simulated using MM11 [13] is displayed in Fig. 9.13. GIDL is dominant for negative values of the gate-to-source voltage, while impact ionization dominates for the conducting transistor.

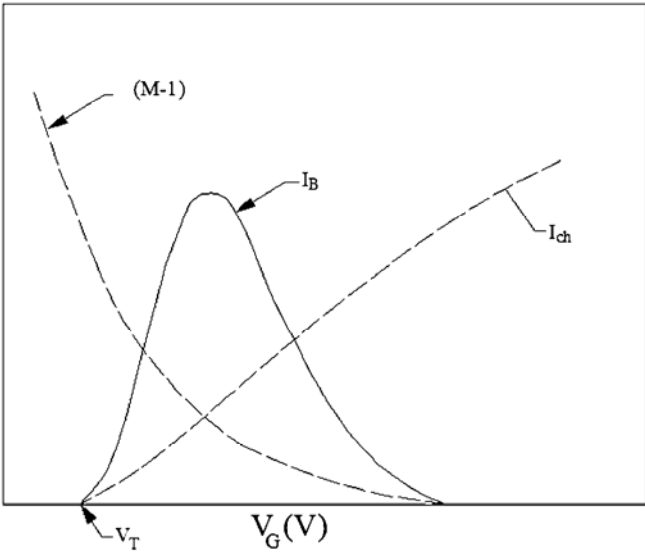


Fig. 9.11 Qualitative representation of substrate current vs. gate voltage V_G for a fixed drain voltage V_D . (After [23].)

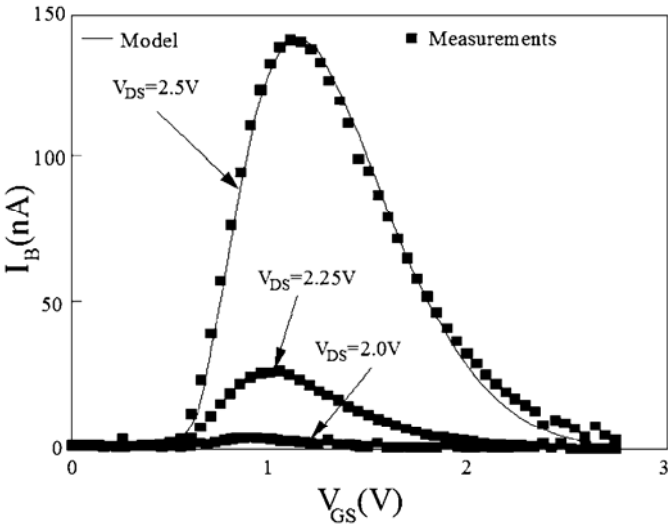


Fig. 9.12 Bulk current I_B as a function of the gate to source bias V_{GS} for different values of the drain-source voltage V_{DS} and $V_{SB}=0V$ (n-MOS, $W/L=10\mu m/0.5\mu m$, $t_{ox}=5.0nm$). (After [13].)

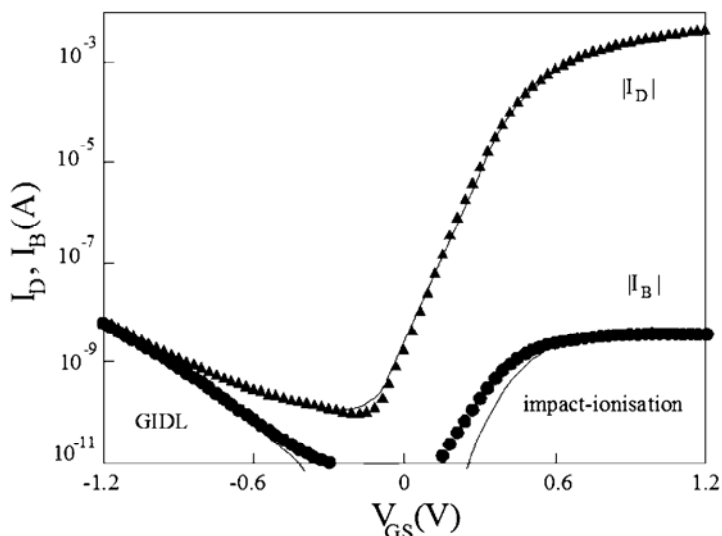


Fig. 9.13 Measured (symbols) and simulated (lines) drain and bulk currents as functions of V_{GS} at high drain bias ($V_{DS}=1.2V$). (After [13].)

9.2.2.2 Breakdown

In most MOS transistors, drain-source breakdown is neither simple junction breakdown nor source-to-drain punchthrough but avalanche-induced breakdown [27]. Junction breakdown can be one of the breakdown mechanisms, especially when the transistor is OFF, but even in this case, the proximity of the gate electrode affects the junction breakdown (see previous discussion on GIDL). When the transistor is ON and the drain voltage is high, impact ionization of electrons in the high electric field close to the drain generates both secondary electrons and holes, as explained before. Holes collected by the substrate contact cause a voltage drop across the substrate resistance that tends to forward bias the source-to-substrate junction. The parasitic lateral npn bipolar transistor can turn on and generate an additional electron current which, in turn, increases the drain current and the number of holes injected into the substrate by avalanche multiplication. The action of the lateral bipolar transistor, together with the current plotted against the drain voltage is illustrated in Fig. 9.14.

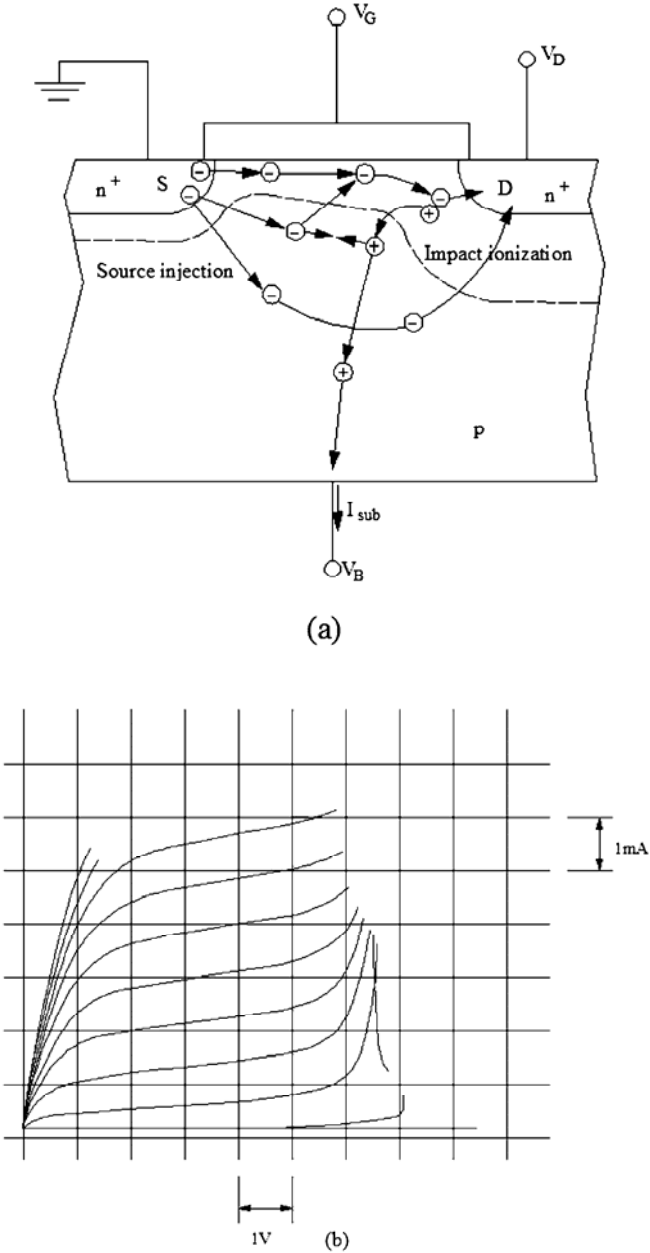


Fig. 9.14 (a) Cross-section of a short-channel MOSFET showing impact ionization currents. (After [27].) (b) Output I-V characteristics.

Another important effect that originates from the substrate current is the so-called substrate current-induced body effect (SCBE) [27]. In fact, when the voltage across the source-to-substrate junction increases due to the substrate current, the amount of depleted charge closer to the source diminishes, giving rise to an increase in the electron charge and, as a result, an increase in the drain current. In the technical literature, this phenomenon is considered to be caused by a reduction in the threshold voltage.

Bibliography

- [1] M. Lenzlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2 ," *J. Appl. Phys.*, vol. 40, no. 1, pp. 278-283, Jan. 1969.
- [2] Michael Shur, *Introduction to Electronic Devices*, John Wiley & Sons, New York, 1996.
- [3] Andreas Gehring and Siegfried Selberherr, "Gate currents in small MOSFETs," presentation at the EDS South-Brazil Chapter Meeting, Campinas, Brazil, 2004.
- [4] R. Clerc, B. de Salvo, G. Ghibaudo, G. Reimbold, and G. Pananakakis, "Electrical characterization and modeling of MOS structures with an ultra-thin oxide," *Solid-State Electron.*, vol. 46, no. 3, pp. 407-416, Mar. 2002.
- [5] Farhan Rana, Sandip Tiwari, and D. A. Buchanan, "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides," *Appl. Phys. Lett.*, vol. 69, no. 8, pp. 1104-1106, 1996.
- [6] Leonard F. Register, Elyse Rosenbaum, and Kevin Yang, "Analytic model for direct tunneling current in polycrystalline silicon-gate metal-oxide-semiconductor devices," *Appl. Phys. Lett.*, vol. 74, no. 3, pp. 457-459, 1999.
- [7] Klaus F. Schuegraf and Chenming Hu, "Hole injection SiO_2 breakdown model for very low voltage lifetime extrapolation," *IEEE Trans. Electron Devices*, vol. 41, no. 5, pp. 761-767, May 1994.
- [8] K. M. Cao, W.-C. Lee, W. Liu, X. Jin, P. Su, S. K. H. Fung, J. X. An, B. Yu, and C. Hu, "BSIM4 gate leakage model including source-drain partition," *IEDM Tech. Dig.*, 2000, pp. 815-818.
- [9] R. Clerc, P. O'Sullivan, K. G. McCarty, G. Ghibaudo, G. Pananakakis, A. Mathewson, "A physical compact model for direct tunneling from NMOS inversion layers," *Solid-State Electron.*, vol. 45, no. 10, pp. 1705-1716, Oct. 2001.
- [10] Z. A. Weinberg, "On tunneling in metal-oxide-silicon structures," *J. Appl. Phys.*, vol. 53, no. 7, pp. 5052-5056, July 1982.
- [11] Frank Stern, "Quantum properties of surface space-charge layers," *CRC Crit. Rev. Solid State Sci.*, pp. 499-514, 1974.
- [12] Z. A. Weinberg, "Tunneling of electrons from Si into thermally grown SiO_2 ," *Solid-State Electron.*, vol. 20, no. 1, pp. 11-18, Jan. 1977.

- [13] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "Physical Background of MOS Model 11," Nat. Lab. Unclassified Report 2003/00239. [Online]. Available: http://www.semiconductors.philips.com/Philips_Models/
- [14] R. Clerc, A. Spinelli, G. Ghibaudo, and G. Pananakakis, "Theory of direct tunneling current in metal-oxide-semiconductor structures," *J. Appl. Phys.*, vol. 91, no. 3, pp. 1400-1409, Feb. 2002.
- [15] Jin Cai and Chih-Tang Sah, "Gate tunneling currents in ultrathin oxide metal-oxide-silicon transistors," *J. Appl. Phys.*, vol. 89, no. 4, pp. 2272-2285, Feb. 2001.
- [16] Wei-Kai Shih, Rafael Rios, Paul Packan, Kaizad Mistry, Tracy Abbott, "A general partition scheme for gate leakage current suitable for MOSFET compact models," *IEDM Tech. Dig.*, 2001, pp. 293-296.
- [17] R. van Langevelde, A. J. Scholten, R. Duffy, F. N. Cubaynes, M. J. Knitel, and D. B. M. Klaassen, "Gate current: modeling, ΔL extraction and impact on RF performance," *IEDM Tech. Dig.*, 2001, pp. 289-292.
- [18] J. Chen, T. Y. Chan, I. C. Chen, P. K. Ko, and Chenming Hu, "Subbreakdown drain leakage current in MOSFET," *IEEE Electron Device Lett.*, vol. 8, no. 11, pp. 515-517, Nov. 1987.
- [19] Tetsuo Endoh, Riichiroh Shirota, Masaki Momodomi, and Fujio Masuoka, "An accurate model of subbreakdown due to band-to-band tunneling and some applications," *IEEE Trans. Electron Devices*, vol. 37, no. 1, pp. 290-296, Jan. 1990.
- [20] Ja-Hao Chen, Shyh-Chyi Wong, and Yeong-Her Wang, "An analytic three-terminal band-to-band tunneling model on GIDL in MOSFET," *IEEE Trans. Electron Devices*, vol. 48, no. 7, pp. 1400-1405, July 2001.
- [21] S. M. Sze, *Physics of Semiconductor Devices*, Wiley-Interscience, New York, 1969.
- [22] C. A. Lee, R. A. Logan, R. L. Batdorf, J. J. Kleimack, and W. Wiegmann, "Ionization rates of holes and electrons in silicon," *Phys. Rev.*, vol. 134, no. 3A, pp. A761-773, 1964.
- [23] Y. A. El-Mansy and D. M. Caughey, "Modeling weak avalanche multiplication currents in IGFETS and SOS transistors for CAD," *IEDM Tech. Dig.*, 1975, pp. 31-34.
- [24] Ronald R. Troutman, "Low-level avalanche multiplication in IGFET's," *IEEE Trans. Electron Devices*, vol. 23, no. 4, pp. 419-425, Apr. 1976.
- [25] J. W. Slotboom, G. Streutker, G. J. T. Davids, and P. B. Hartog, "Surface impact ionization in silicon devices," *IEDM Tech. Dig.*, 1987, pp. 494-497.
- [26] Narain Arora, *MOSFET Models for VLSI Circuit Simulation Theory and Practice*, Springer-Verlag, Wien, 1993.
- [27] Fu-Chieh Hsu, Ping-Keung Ko, Simon Tam, Chenming Hu, and Richard S. Muller, "An analytical breakdown model for short-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 29, no. 11, pp. 1735-1740, Nov. 1982.
- [28] Chang-Hoon Choi, Ki-Young Nam, Zhiping Yu, and Robert W. Dutton, "Impact of gate direct tunneling current on circuit performance: a simulation study," *IEEE Trans. Electron Devices*, vol. 48, no. 12, pp. 2823-2829, Dec. 2001.

Problems

9.1. Using the current division principle (see problem 3.4), derive equations (9.1.28) and (9.1.29). (Hint: Suppose that an elementary gate current $dI_G(y) = WJ_G(y)dy$ flows into node Y . Calculate the fraction of the current that flows to the drain and source ends).

This page intentionally left blank

Chapter 10

Advanced MOSFET Structures

In the previous chapters we developed the models for the conventional bulk CMOS transistor. This chapter presents, in some detail, different MOSFET structures and their modeling issues. Firstly, we summarize some (deep) submicron planar MOS structures. We then review the silicon-on-insulator (SOI) MOSFET. Finally, we introduce some innovative MOS devices that may replace the conventional MOS transistor in deeply scaled CMOS technologies. It is shown that the charge based model, with some minor variations, is appropriate for all the different MOS structures herein.

10.1 Introduction

Physics places constraints on the conventional bulk MOS transistor for deeply scaled technologies. To keep pace with the MOSFET scaling predicted by the International Technology Roadmap for Semiconductors [1], new materials and devices have been and will have to be introduced. Shrinking MOS transistors beyond the 50 nm technology node poses some challenges for the semiconductor industry such as (i) quantum-mechanical tunneling of carriers through the thin gate oxide, from source to drain, and from drain to the body of the MOSFET; (ii) control of the dopant number fluctuations; (iii) increased short-channel effects such as DIBL, and (iv) a reduction in the on-off current ratio [2], [3]. To overcome these challenges, modifications to the classical bulk MOS transistor have been proposed. We will focus the next sections on modeling such non-classical CMOS devices as fully depleted SOI

transistors, surrounding-gate transistors, and multiple-gate FETs. Before presenting these non-classical CMOS devices, we will discuss some technological modifications to the conventional bulk CMOS transistor that have allowed it to be scaled to channel lengths below 100 nm.

10.2 Deep submicron planar MOS transistor structures

In the previous chapters, the model for conventional planar bulk CMOS transistors was described. In this section, we introduce some limitations of conventional MOS transistors for sub-100 nm devices and discuss modeling aspects associated with some of the improvements in the transistor structure.

When the MOSFET channel length is scaled down, the vertical dimensions, *i.e.*, the gate oxide thickness and the source-drain junction depth must be scaled down as well, in order to keep the short-channel effects (SCE) within acceptable limits. Ideal scaling is, for several reasons, not always possible and SCE can be worse in the scaled-down technology. One of the main SCE is the reduction in the threshold voltage with decreasing channel length. Threshold voltage reduction causes the off-current of an MOS transistor to increase significantly, thus giving rise to higher static power dissipation. To comply with the requirements of technology scaling, devices have been designed with more complex doping profiles in an effort to maintain long channel behavior at short channel lengths.

To illustrate some differences with respect to the idealized CMOS technology, in which the channel doping profile is uniform, the schematic cross section of a 0.1- μm CMOS technology is shown in Fig. 10.1 [3]. The gate oxide thickness is 35-Å. Shallow p+ (or n+) source-drain extensions are used in conjunction with deeper p+ (or n+) source-drain regions implanted after thick oxide spacers. A medium-dose, counter-doping implant (halo) is obtained with the extension implant to increase the channel doping level to limit short-channel effects. The resulting impurity concentration near the source and drain is non-uniform and the average doping concentration in the channel is a function of the channel length. As a result, the SCE are reduced due to

a more effective control of the channel by the gate terminal. For a more detailed description of the technology associated with Fig. 10.1, the reader is referred to [4].

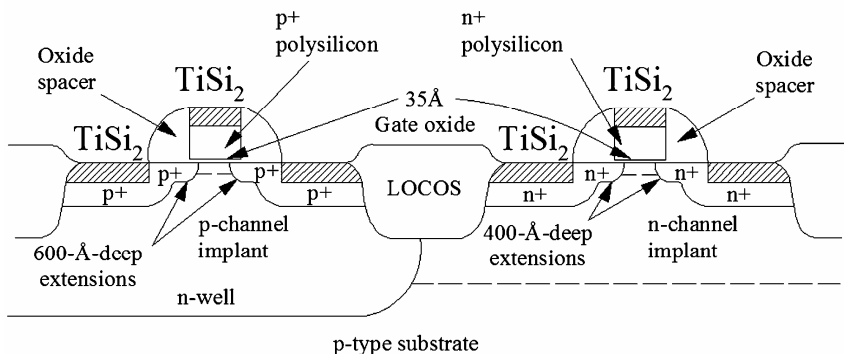


Fig. 10.1 Schematic cross section of 0.1- μm CMOS devices. (After [3].)

Current CMOS technologies make use of halo or pocket implants to reduce short-channel effects [5], [6], [7], [8]. These processes, which enable gate lengths in the sub-100 nm region, result in non-uniform profiles such as the pocket-implanted profile in Fig. 10.2 [8] or the halo implant in Fig. 10.3 [9].

The transistor resulting from typical halo processes can be modeled as a series connection of three transistors formed by a central device and two edge devices, whose characteristics differ from the central device, as shown in Fig. 10.4. This partition of the transistor for calculation and simulation purposes, even though physically sound, has the following drawbacks: (i) two additional nodes are required in the three-transistor counterpart; (ii) the parameters associated with the center and the edge transistors must be characterized, namely, length, threshold voltage, and mobility.

A second approach [5], [8] used to model the MOSFET fabricated with the halo/pocket process is to divide the transistor as illustrated in Fig. 10.4 and then find an “equivalent” transistor with a uniform doping that is dependent on the doping profile of the halo or pocket-implanted MOS transistor. Even though the calculation of the uniform doping of the

equivalent transistor is slightly different in references [5] and [8], for a first-order calculation one can think of the equivalent transistor as having a uniform doping equal to the average doping of the halo/pocket-implanted transistor. Using the average doping profile for the calculation of the uniform doping in the equivalent transistor, the authors of [8] found that the deviation between measured and modeled threshold voltages for both p- and n-channel transistors does not exceed 5 mV in 0.1 μm CMOS technologies, thus confirming the model's suitability for the threshold voltage.

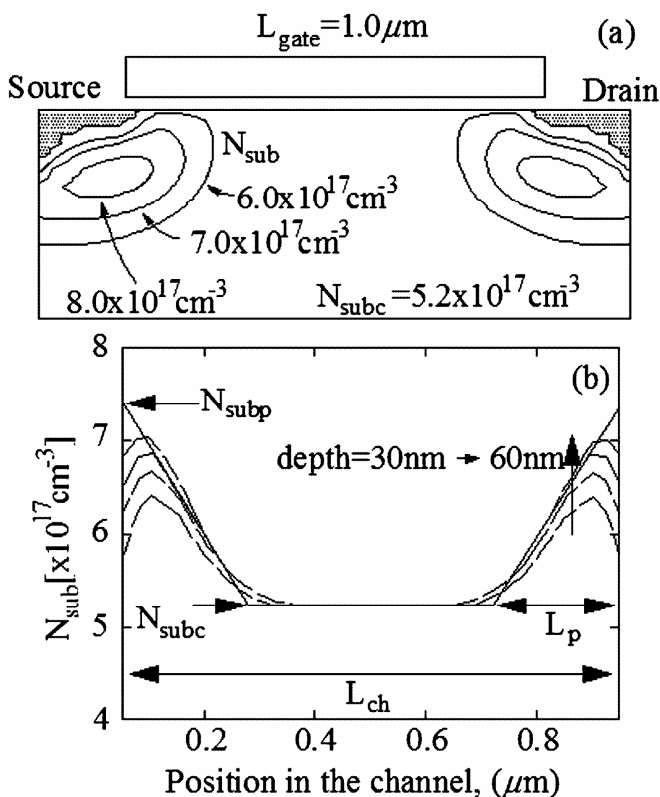


Fig. 10.2 Pocket profiles extracted from measurements (a) 2-D extraction from process simulator (b) Extracted model for circuit simulation (thick solid line). The dashed curves are the results from the process simulator for various depths. (After [8].)

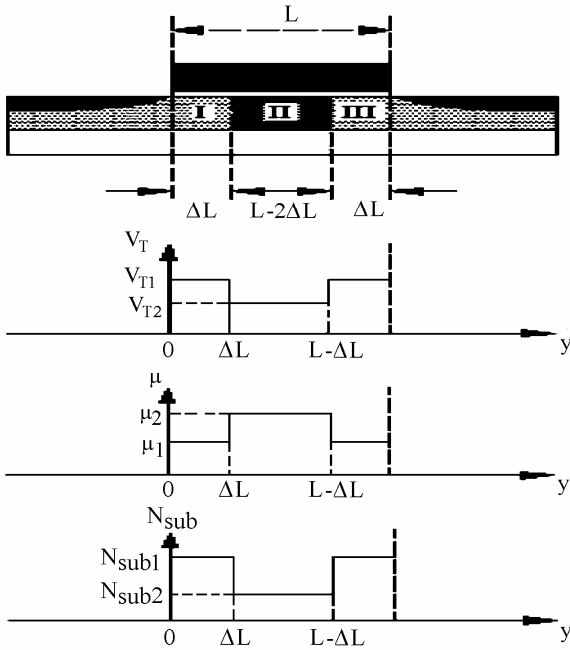


Fig. 10.3 Channel impurity profile of a halo-implanted MOSFET. (After [9].)

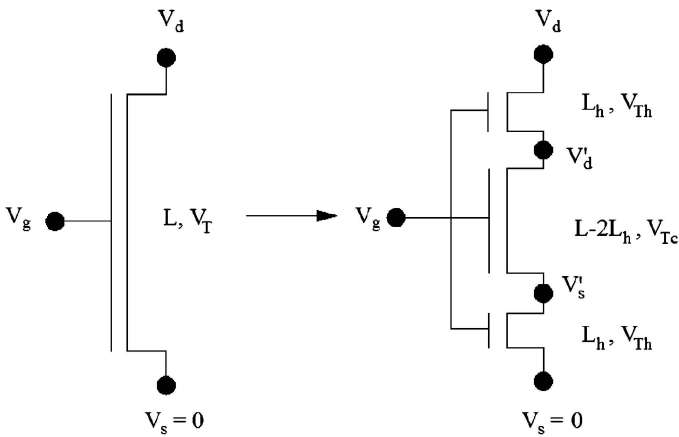


Fig. 10.4 Schematic representation of the MOSFET fabricated with halo process, formed by a center device and two edge devices. (After [5].)

Figure 10.5 displays the measured and modeled threshold voltages of pocket-implanted transistors in a 0.1 μm CMOS technology [8] for several channel lengths. Even though the modeled threshold voltage fits the experimental results very well, the same cannot be said of the current dependence on the voltages, due to mobility differences in the center and edge regions as well as relative differences in the weights of the regions in weak and strong inversion.

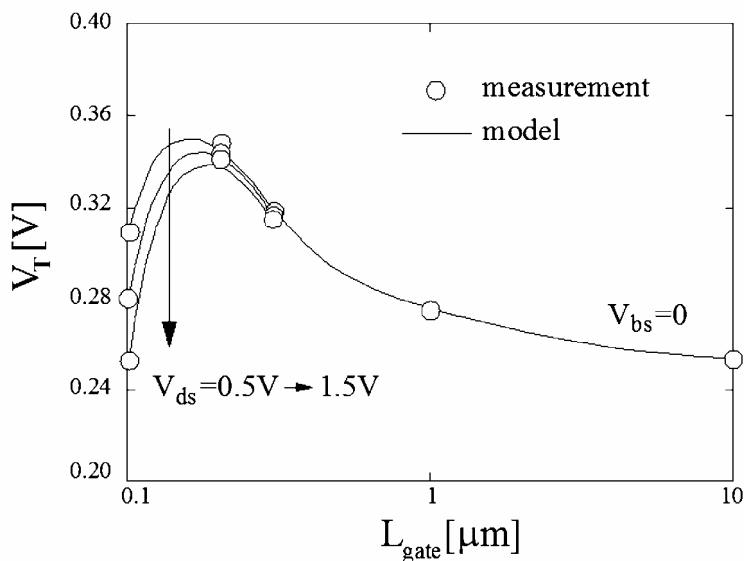


Fig. 10.5 Threshold voltage of MOSFETs for a pocket-implanted 0.1 μm CMOS technology. (After [8].)

10.3 Silicon-on-insulator (SOI) CMOS transistors

Since the 1970s, SOI CMOS has been identified as one alternative for increasing the transistor performance over that offered by conventional bulk CMOS technology [10]. Prior to the 1990s, the barriers to its widespread use were poor SOI material quality and device design, and the steady progress in bulk CMOS performance through scaling [10]. A huge investment by some companies [10] has led to SOI technologies

mature enough to enable the launch of products to compete with those fabricated with bulk CMOS technologies.

Figure 10.6 shows a schematic cross-section of a pair of CMOS transistors in SOI technology, built on very thin body regions. The transistors are separated from the silicon substrate by a relatively thick layer of buried silicon oxide, which electrically isolates the devices from the substrate and from each other [12].

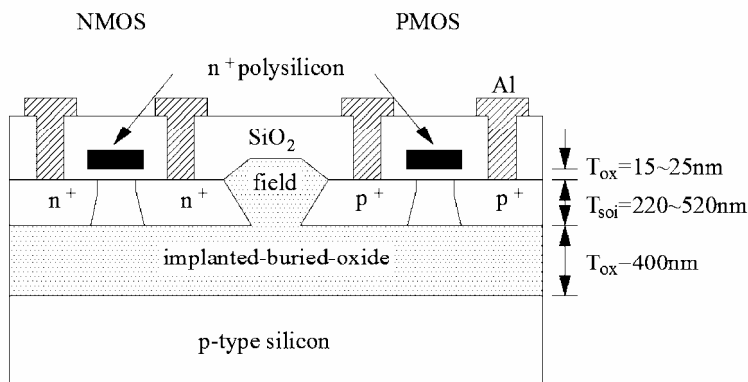


Fig. 10.6 Schematic cross-section of SOI CMOS transistor pair. (After [11].)

The advantages of SOI transistors over bulk transistors are increased radiation hardness, reduction in parasitic capacitances, higher packing density, latch-up-free operation, and possibility of 3-D integration.

Disadvantages of SOI transistors compared with bulk transistors are the floating-body effects and self-heating. This latter factor is usually not of major concern in bulk CMOS technologies since the underlying substrate material has good thermal conductivity. On the other hand, in SOI technologies, the existence of an isolating buried oxide layer of very low thermal conductivity between the devices and the silicon substrate can give rise to important temperature gradients even at moderate power levels [13]. The resultant temperature gradients can adversely affect the circuit performance.

SOI transistors can operate as partially depleted (PD) or fully depleted (FD) devices, depending upon the silicon film thickness and doping. If the silicon film is sufficiently thick, it is never completely

depleted, and as a consequence there is no electrostatic coupling between the front and back interfaces of the semiconductor film. The conduction in the front and back channels is modeled by conventional bulk MOS theory, but the (local) substrate potential exhibits the so-called floating-body effect. The floating-body effect is caused by storage of carriers of the same type as the body, which are generated by impact ionization near the drain region [12]. In an n-channel transistor, the storage of carriers in the neutral body increases the body-to-source voltage which, in turn, results in a reduction in the threshold voltage or, equivalently, in a sharp increase in the output conductance of the device, usually known as the kink effect. Under dynamic conditions, the body-to-source voltage depends not only on the instantaneous voltage but also on the previous electrical “history” of the device [10].

One method used to overcome floating-body effects is to fabricate FD SOI devices. In FD transistors, the silicon film thickness is thinner than the channel depletion width [10], [12] and, therefore, the body capacitance is close to zero. Hence, the value of the subthreshold slope can be theoretically very close to the ideal value of 60 mV/dec at 300 K which, however, occurs only in long-channel devices. In fact, it has been shown that for a given channel length, FD SOI devices exhibit increased SCE compared to PD SOI devices unless the silicon film thickness becomes much smaller than the junction depth [10], [14]. To obtain reduced SCE in FD SOI transistors, the silicon film thickness must be considerably smaller than the junction depth of a comparable bulk technology because of an additional phenomenon due to charge-sharing through the thick buried oxide [14]. Even though the control of the threshold voltage in FD transistors is not as simple as in PD transistors [10], FD transistors are not prone to the kink effect. In FD transistors, the conventional bulk MOS transistor model is not valid because of the electrostatic coupling between the two gates. For this reason we will develop the fundamental equations for charge and current of the FD transistor.

Figure 10.7 illustrates the schematic cross section of an n-channel SOI transistor. To derive the basic equations for the SOI transistor, we initially assume it to be a long-channel device, *i.e.*, the y -component of

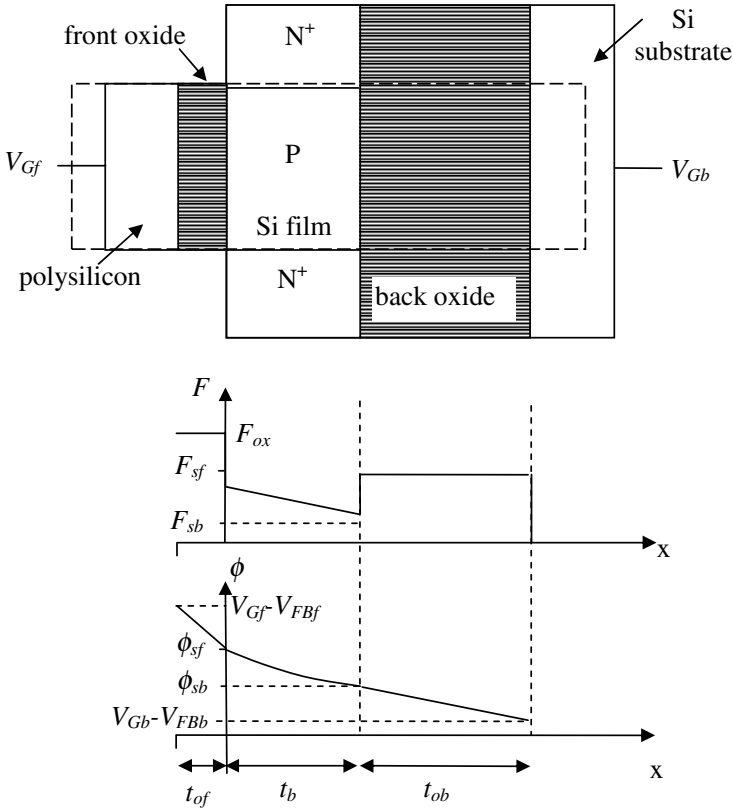


Fig. 10.7 Cross-section of an SOI MOSFET structure, electric field, and electrostatic potential along the SOI structure.

the electric field is negligible compared to the x-component. Additionally, the uniformly doped silicon film is assumed to be fully depleted, from the front surface to the back surface. Taking into account the depletion approximation together with the 1-D approximation of the transistor, we can derive the electric field and electrostatic potential at a given y-coordinate inside the enclosed dashed box, as shown in

Fig. 10.7. The front ϕ_{sf} and back ϕ_{sb} surface potentials are the band bending from a (hypothetical) neutral film body point to the respective surface. The continuity of the electrical displacement at the front and back interfaces ($\epsilon_{ox} F_{ox} = \epsilon_s F_s$) yields [15], [16], [17]

$$V_{Gf} - V_{FBf} = \phi_{sf} + \frac{\epsilon_s F_{sf}}{C_{of}}, \quad (10.3.1)$$

$$V_{Gb} - V_{FBb} = \phi_{sb} - \frac{\epsilon_s F_{sb}}{C_{ob}}, \quad (10.3.2)$$

where F_{sf} and F_{sb} are the front- and back-surface electric fields, C_{of}^1 and C_{ob} are the front- and back-gate oxide capacitances per unit area, and V_{FBf} and V_{FBb} are the front- and back-gate flat-band voltages, respectively. Equations (10.3.1) and (10.3.2) are analogous to the bulk MOS equation (2.3.34) for potential balance, but for an FD device we must solve a system of equations to obtain the charges because of the electrostatic coupling between the gates.

Using the charge sheet approximation for both the front and back interfaces we can write the charge density along the silicon film as

$$\rho = -qN_A + Q_{nf}\delta(x) + Q_{nb}\delta(x - t_b), \quad (10.3.3)$$

where $\delta(\bullet)$ is the Dirac impulse, N_A is the film doping, Q_{nf} is the electron charge density at the front interface, and Q_{nb} is the electron or hole charge density at the back interface. Both Q_{nf} and Q_{nb} have been approximated as charge sheets [16].

Now using the Poisson equation in the silicon film, and the relationship between electric field and potential, yields

$$F = F_{sf} \text{ for } x = 0, \quad (10.3.4)$$

$$F = F_{sf} + \frac{Q_{nf}}{\epsilon_s} - \frac{qN_A}{\epsilon_s} x \text{ for } 0 < x < t_b, \quad (10.3.5)$$

¹ In this chapter, C and Q represent capacitance and charges per unit area. We decided to keep the symbols as usually represented in the technical literature on SOI and multiple-gate devices rather than using C' and Q' of the previous chapters for capacitance and charge densities.

$$F = F_{sb} = F_{sf} + \frac{Q_{nf}}{\epsilon_s} + \frac{Q_{nb}}{\epsilon_s} - \frac{qN_A}{\epsilon_s} t_b \text{ for } x = t_b, \quad (10.3.6)$$

$$\phi = \phi_{sf} - (F_{sf} + \frac{Q_{nf}}{\epsilon_s})x + \frac{qN_A}{2\epsilon_s} x^2 \text{ for } 0 < x < t_b, \quad (10.3.7)$$

$$\phi_{sb} = \phi_{sf} - (F_{sf} + \frac{Q_{nf}}{\epsilon_s})t_b + \frac{qN_A}{2\epsilon_s} t_b^2 \text{ for } x = t_b. \quad (10.3.8)$$

From (10.3.8) it follows that

$$\frac{\phi_{sf} - \phi_{sb}}{t_b} = F_{sf} + \frac{Q_{nf}}{\epsilon_s} - \frac{qN_A t_b}{2\epsilon_s}. \quad (10.3.9)$$

Substituting the value of F_{sf} from (10.3.1) into (10.3.9) we obtain

$$Q_{nf} = -C_{of} \left[V_{Gf} - V_{FBf} - \left(1 + \frac{C_b}{C_{of}} \right) \phi_{sf} + \frac{C_b}{C_{of}} \phi_{sb} + \frac{Q_b}{2C_{of}} \right], \quad (10.3.10)$$

where $Q_b = -qN_A t_b$ is the depletion charge density of the silicon film and $C_b = \epsilon_s / t_b$ is the semiconductor film capacitance per unit area. Thus, according to the charge-sheet model, for a long-channel FD device, the relation between the inversion charge density and the surface potentials is linear.

In a similar way we obtain

$$Q_{nb} = -C_{ob} \left[V_{Gb} - V_{FBb} - \left(1 + \frac{C_b}{C_{ob}} \right) \phi_{sb} + \frac{C_b}{C_{ob}} \phi_{sf} + \frac{Q_b}{2C_{ob}} \right]. \quad (10.3.11)$$

The validity of analytical expressions (10.3.10) and (10.3.11), which describe the charge densities at the front and back channels assuming the charge-sheet model, has been verified in reference [18] through comparison with numerical analysis.

To simplify the two-dimensional problem of the non-equilibrium long-channel SOI MOS transistor to a one-dimensional problem, we use the gradual channel and the quasi-equilibrium approximations [17]. The quasi-equilibrium approximation, which is illustrated in Fig. 10.8,

assumes that the quasi-Fermi levels are constant across the silicon film [17]. The quasi-equilibrium approximation is used to analyze the one-dimensional SOI MOSFET for the case in which the front surface is inverted and the back surface is accumulated or depleted. In other words, the analysis will be restricted to cases in which the back surface does not contribute to the device current.

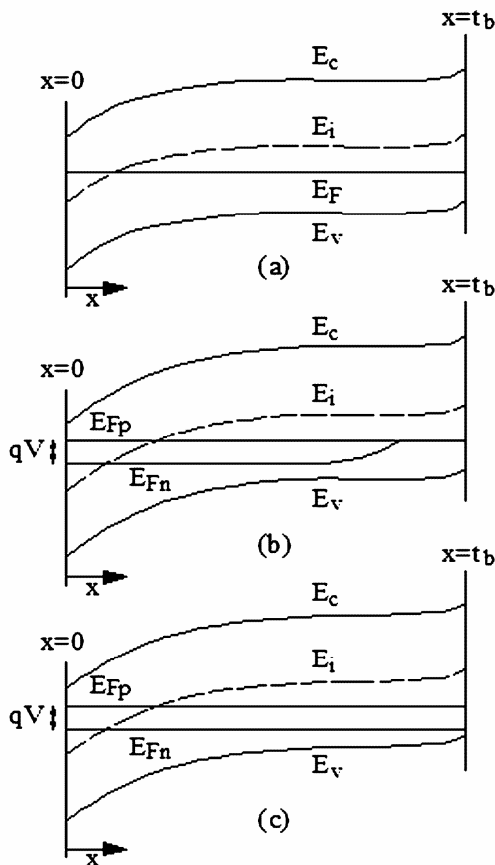


Fig. 10.8 Diagrams of electron energy-bands perpendicular to the channel in a specific y -coordinate: (a) thermal equilibrium ($V_D = 0$), (b) non thermal equilibrium ($V_D > 0$), and (c) non thermal equilibrium ($V_D > 0$) using the quasi-equilibrium approximation. (After [17].) $V = (E_{Fp} - E_{Fn}) / q$ represents the splitting of the quasi-Fermi levels along the channel.

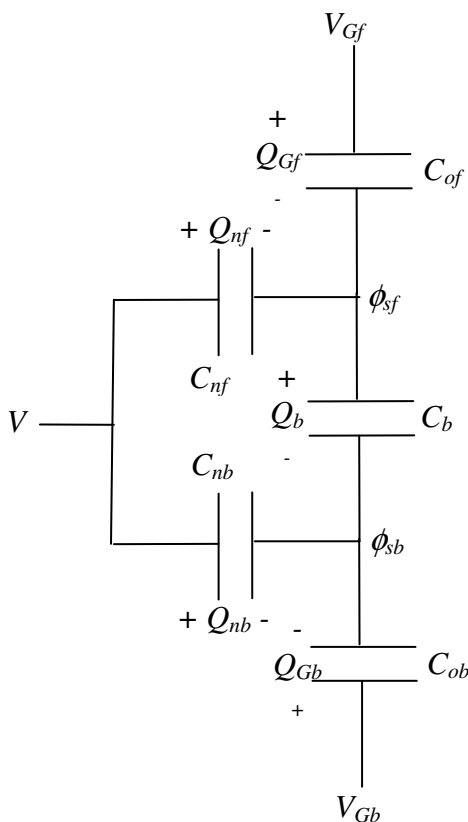


Fig. 10.9 Capacitive model of the SOI transistor using the quasi-equilibrium approximation for the back surface either in accumulation or in depletion.

By analogy with the capacitive model of the three-terminal bulk MOS transistor, the capacitive model of the SOI transistor is shown in Fig. 10.9. Adopting the same approximation of the inversion capacitance $C_{nf} = -Q_{nf}/\phi_t$ as that adopted for the bulk transistor, in the following we will derive the basic equations for finding the inversion charge density in terms of the terminal voltages. In particular, we assume that the back surface is in depletion, a condition that allows us to determine very simple expressions for the inversion charge at the front surface and to gain an insight into the influence of the physical parameters on the device characteristics.

Assuming that the back surface is depleted, we can write $Q_{nb}=0$. The substitution of $Q_{nb}=0$ into (10.3.11) leads to the following relationship between the back- and front-surface potentials

$$\phi_{sb} = \frac{1}{(1 + C_b/C_{ob})} \left(V_{Gb} - V_{FBb} + \frac{C_b}{C_{ob}} \phi_{sf} + \frac{Q_b}{2C_{ob}} \right). \quad (10.3.12)$$

Now we substitute the value of ϕ_{sb} from (10.3.12) into (10.3.10), which gives

$$\begin{aligned} Q_{nf} = -C_{of} \left[V_{Gf} - V_{FBf} - \left(1 + \frac{C_{bb}}{C_{of}} \right) \phi_{sf} \right. \\ \left. + \frac{C_{bb}}{C_{of}} \left(V_{Gb} - V_{FBb} + \frac{Q_b}{2C_{ob}} \right) + \frac{Q_b}{2C_{of}} \right] \end{aligned} \quad (10.3.13)$$

where $C_{bb} = C_{ob}C_b/(C_{ob} + C_b)$ is the series association of the film capacitance and the back-oxide capacitance.

According to (10.3.13), for constant front and back-gate voltages, the variation in the inversion charge density dQ_{nf} in the channel follows the variation in the front-surface potential according to

$$dQ_{nf} = (C_{of} + C_{bb}) d\phi_{sf}. \quad (10.3.14)$$

In an FD SOI MOSFET, the capacitive coefficient in (10.3.14), *i.e.*, the derivative of the inversion charge density with respect to the surface potential is constant for any inversion level [19]. In practice, usually $C_{of} \gg C_{bb}$.

Recalling that the drift and diffusion components of the channel current are

$$I_D = I_{drift} + I_{diff} = -\mu_n W Q_{nf} \frac{d\phi_{sf}}{dy} + \mu_n W \phi_t \frac{dQ_{nf}}{dy} \quad (10.3.15)$$

we can determine, similarly to the case of the bulk MOS transistor, the channel pinch-off charge density Q_{fp} as being that at which the drift and diffusion components of the current are equal [20]. Then, using (10.3.14) and (10.3.15) we find that

$$Q_{fp} = -(C_{of} + C_{bb})\phi_t = -n_f C_{of} \phi_t, \quad (10.3.16)$$

where $n_f = 1 + C_{bb}/C_{of} \cong 1$ is the (front-gate) slope factor. Now, inserting the value of Q_{fp} from (10.3.16) into (10.3.13), we determine the value of the front-surface potential ϕ_{sp} at which the inversion charge density equals Q_{fp} .

$$\begin{aligned} \phi_{sp} = & \frac{C_{of}}{C_{of} + C_{bb}} (V_{Gf} - V_{FBf}) + \frac{C_{bb}}{C_{of} + C_{bb}} (V_{Gb} - V_{FBb}) \\ & + \frac{C_{bb} + C_{ob}}{C_{of} + C_{bb}} \frac{Q_b}{2C_{ob}} - \phi_t. \end{aligned} \quad (10.3.17)$$

Note that, for the usual case of $C_{of} \gg C_{bb}$, the effect of the back gate on the surface potential at pinch-off is significantly smaller than the effect of the front gate.

To determine the charge-to-voltage relationship we use the capacitive model of the SOI transistor represented in Fig. 10.9. For constant front- and back-gate voltages, and for a depleted back-surface ($Q_{nb}=0$), we can write

$$\frac{dQ_{nf}}{dV} = \frac{C_{nf} (C_{of} + C_{bb})}{C_{of} + C_{bb} + C_{nf}}. \quad (10.3.18)$$

The substitution of $C_{nf} = -Q_{nf}/\phi_t$ into (10.3.18) yields

$$\left(\frac{-\phi_t}{Q_{nf}} + \frac{1}{C_{of} + C_{bb}} \right) dQ_{nf} = dV. \quad (10.3.19)$$

The integration of (10.3.19) from the pinch-off voltage in the channel to a generic voltage leads to

$$\frac{V_P - V}{\phi_t} = \frac{Q_{nf}}{Q_{fp}} - 1 + \ln \frac{Q_{nf}}{Q_{fp}}, \quad (10.3.20)$$

which is the unified charge control model (UCCM) adapted to SOI transistors [19]. As shown in the next section the approximate pinch-off voltage is given by

$$V_p \cong \phi_{sp} - \phi_t \ln \frac{C_{of}}{C_b} - 2\phi_F \quad (10.3.21)$$

and the equilibrium threshold voltage by

$$V_{T0} \cong V_{FBf} - \frac{Q_b}{C_{of}} + 2\phi_F + \phi_t \left(1 + \ln \frac{C_{of}}{C_b} \right). \quad (10.3.22)$$

Finally, using the approximate linear relationship of the pinch-off in terms of the gate voltage, UCCM is written as

$$\frac{V_{Gf} - V_{T0} - n_f V}{n_f \phi_t} = \frac{Q_{nf}}{Q_{fp}} - 1 + \ln \frac{Q_{nf}}{Q_{fp}}, \quad (10.3.23)$$

where the values of V_{T0} , n_f and Q_{fp} are given in (10.3.22) and (10.3.16).

Once we have determined the relationship between inversion charge and external voltages, we can write the drain current as

$$I_D = \mu_n \frac{W}{L} \phi_t \left[\frac{Q_{nf,s}^2 - Q_{nf,d}^2}{2n_f C_{of} \phi_t} - (Q_{nf,s} - Q_{nf,d}) \right]. \quad (10.3.24)$$

Based on equations (10.3.23) and (10.3.24), the charges and capacitances of the long-channel SOI transistor can be calculated as shown in [19]. Expressions for the computation of small-geometry effects are also presented in [19].

10.3.1 Pinch-off voltage calculation ²

To calculate V_p , we use the “refined depletion approximation” [17], [21], which can be described as follows. Firstly, as in section 2.2.2 we write Poisson’s equation as

$$\frac{dF^2}{d\phi} = -\frac{2\rho}{\epsilon_s} \quad (10.3.25)$$

where

² This section can be omitted in an introductory study, without loss of continuity.

$$\rho = q \left[p - n - (p_0 - n_0) \right] = qp_0 \left[e^{\frac{-\phi}{\phi_i}} - 1 - \frac{n_0}{p_0} \left(e^{\frac{\phi-V}{\phi_i}} - 1 \right) \right], \quad (10.3.26)$$

n_0 and p_0 being the electron and hole equilibrium concentrations. Integrating (10.3.25) from ϕ_{sf} to ϕ we obtain

$$F^2 - F_{sf}^2 = G^2(\phi, V) - G^2(\phi_{sf}, V) \quad (10.3.27)$$

where

$$G^2(\phi, V) = -\frac{2}{\epsilon_s} \int_0^\phi \rho(\phi, V) d\phi. \quad (10.3.28)$$

The refined depletion approximation consists of neglecting the contribution of the minority carriers to the charge density, *i.e.*

$$\rho_{da} = q[p - p_0] = qp_0 \left[e^{\frac{-\phi}{\phi_i}} - 1 \right]. \quad (10.3.29)$$

Similarly to (10.3.28), we define the following function

$$G_{da}^2(\phi) = -\frac{2}{\epsilon_s} \int_0^\phi \rho_{da}(\phi) d\phi. \quad (10.3.30)$$

As presented and explained in [17] and [21], the key approximation

$$F_{sf}^2 - G^2(\phi_{sf}, V) \approx F_{sfda}^2 - G_{da}^2(\phi_{sf}) \quad (10.3.31)$$

can be used to calculate the relationship between the front-surface field F_{sf} and the depletion electric field F_{sfda} at an infinitesimally small depth in the silicon film. Now, using again Gauss' law we can calculate the approximate electron charge density Q_{nfa} as

$$Q_{nfa} = \epsilon_s (F_{sfda} - F_{sf}). \quad (10.3.32)$$

Expression (10.3.31), along with the solutions to the integrals in (10.3.28) and (10.3.30) and expressions (10.3.31) and (10.3.32), allows writing the expression of the inversion charge density in the front-channel as

$$Q_{nfda} = \left\{ C_{of}^2 [V_{Gf} - V_{FBf} - \phi_{sf}]^2 - 2kTn_0\epsilon_s \right. \\ \left. \cdot \left[e^{\frac{V}{\phi_t}} \left(e^{\frac{\phi_{sf}}{\phi_t}} - 1 \right) - \frac{\phi_{sf}}{\phi_t} \right] \right\}^{1/2} - C_{of} [V_{Gf} - V_{FBf} - \phi_{sf}]. \quad (10.3.33)$$

To calculate the pinch-off voltage, we use the definition of the pinch-off charge together with (10.3.33), which gives

$$Q_{fp} = Q_{nfda}(\phi_{sf} = \phi_{sP}, V = V_p) = -(C_{of} + C_{bb})\phi_t \\ = C_{of}(V_{Gf} - V_{FBf} - \phi_{sP}) \left\{ \left[1 - \frac{2kTn_0\epsilon_s}{C_{of}^2 (V_{Gf} - V_{FBf} - \phi_{sP})^2} \left(e^{\frac{\phi_{sP} - V}{\phi_t}} - \frac{\phi_{sP}}{\phi_t} \right) \right]^{1/2} - 1 \right\}. \quad (10.3.34)$$

The approximation $\phi_{sP}/\phi_t \gg 1$ has been used in (10.3.34). The value of ϕ_{sP} in (10.3.17) allows the calculation of the term

$$V_{Gf} - V_{FBf} - \phi_{sP} = \frac{C_{bb}}{C_{of} + C_{bb}} (V_{Gf} - V_{FBf} - V_{Gb} + V_{FBb}) \\ - \frac{C_{bb} + C_{ob}}{C_{of} + C_{bb}} \frac{Q_b}{2C_{ob}} + \phi_t \cong \frac{C_{bb}}{C_{of}} (V_{Gf} - V_{FBf} - V_{Gb} + V_{FBb}) - \frac{Q_b}{C_{of}} + \phi_t. \quad (10.3.35)$$

In (10.3.35), the approximations $C_{of} \gg C_{ob}$ and $C_b \gg C_{ob}$ have been assumed. The term in (10.3.35) can be further simplified for usual cases, where the following approximation holds:

$$\frac{C_{bb}}{C_{of}} (V_{Gf} - V_{FBf} - V_{Gb} + V_{FBb}) - \frac{Q_b}{C_{of}} + \phi_t \cong -\frac{Q_b}{C_{of}}. \quad (10.3.36)$$

Note that the term in braces in (10.3.34) is multiplied by a factor approximately equal to the film charge. Since the pinch-off charge is approximately $-C_{of}\phi_t$ and $-Q_b \gg C_{of}\phi_t$, then the term in braces is much less than one. Thus, using

$$(1-x)^{1/2} \cong 1 - x/2 \quad \text{for } x \ll 1$$

(10.3.34) becomes

$$-C_{of}\phi_t \simeq \frac{kTn_0\epsilon_s}{Q_b} \left(e^{\frac{\phi_{sP}-V}{\phi_t}} - \frac{\phi_{sP}}{\phi_t} \right) = -\frac{n_i^2}{N_A^2} \left(e^{\frac{\phi_{sP}-V}{\phi_t}} - \frac{\phi_{sP}}{\phi_t} \right) C_b\phi_t \quad (10.3.37)$$

Since C_b and C_{of} are of the same order of magnitude and $n_i/N_A \ll 1$, the term in parenthesis is much greater than one, which means the prevalence of the exponential term over the linear term. Recalling that

$$\phi_F = \phi_t \ln \frac{p_0}{n_i} \cong \phi_t \ln \frac{N_A}{n_i}, \quad (10.3.38)$$

the pinch-off voltage determined from (10.3.37) is then given by

$$V_P \cong \phi_{sP} - \phi_t \ln \frac{C_{of}}{C_b} - 2\phi_F, \quad (10.3.39)$$

with ϕ_{sP} calculated from (10.3.17). Recalling that the equilibrium threshold voltage V_{T0} is the (front-) gate voltage for $V_P=0$, one can immediately calculate V_{T0} from (10.3.39) and (10.3.17).

$$\begin{aligned} V_{T0} &\cong V_{FBf} - \frac{C_{bb}}{C_{of}}(V_{Gb} - V_{FBb}) - \frac{Q_b}{C_{of}} + \frac{C_{bb} + C_{of}}{C_{of}} \left[2\phi_F + \phi_t \left(1 + \ln \frac{C_{of}}{C_b} \right) \right] \\ &\simeq V_{FBf} - \frac{Q_b}{C_{of}} + 2\phi_F + \phi_t \left(1 + \ln \frac{C_{of}}{C_b} \right). \end{aligned} \quad (10.3.40)$$

The threshold voltage, as expected, depends on the (front-gate) flat-band voltage and is almost independent of the back-gate bias as long as the back-oxide capacitance is much smaller than the front-oxide and silicon film capacitances. The threshold voltage also depends, to a large extent, on the silicon film charge and on the front oxide capacitance. The expression we have obtained for the threshold voltage is in close agreement with the one presented in [19].

10.4 Undoped surrounding-gate transistors

To obtain reduced short-channel effects in SOI devices in comparison to those in bulk technologies, the silicon film thickness must be smaller than the bulk junction depth [14]. In FD SOI transistors, the close to ideal 60 mV/decade subthreshold slope occurs because the back gate does not have control over the front channel owing to the large buried oxide thickness. On the other hand, the depletion layer in source and drain cannot terminate in the buried oxide but must be counterbalanced by either charge in the back gate or in the channel [3], [14]. Since the buried oxide is thick, a portion of the depletion region associated with the drain junction will terminate in the channel, thus resulting in increased short-channel effects. As a consequence, the lack of back gate control over the channel results in a poor short-channel effect in FD SOI devices. The PD SOI transistor, in turn, does not present the close to ideal subthreshold slope of FD devices and is prone to exhibit the kink effect. Additionally, the random dopant fluctuations in the silicon film of SOI devices can cause severe mismatch between devices.

The scaling down of MOSFETs to sub-50 nm channel length requires the engineering of new device structures. Symmetric/asymmetric double-gate structures, tri-gate transistors, finFETs, and surrounding-gate transistors have been the focus of many research groups due to their improved current drive capability, close to ideal subthreshold slope, improved mobility (due to volume inversion and elimination of impurity scattering) and reduction of statistical dopant fluctuation owing to the use of lightly doped silicon films.

One of the most promising structures for future VLSI devices is the surrounding-gate (SG) MOSFET [22], [23], [24], [25], [26], which is fabricated on a cylindrical pillar of silicon surrounded by the gate electrode. Among the various candidate devices previously mentioned, we have chosen the SG MOSFET for the calculation of the charge density and the current. The reasons for such a choice are two-fold: i) the SG MOSFETs exhibit characteristics, at least with regard to the long-channel device, quite close to those exhibited by multiple-gate

transistors; (ii) the solutions for charge density and drain current are very simple and closely resemble the equations derived for the bulk transistor.

In this section, we present a charge-based model for the undoped (or lightly doped) surrounding-gate MOSFET valid in all transistor operating regions. In particular, we derive expressions for the channel charge and current, which result from the analytical solution to the Poisson equation in a cylindrical structure. The expressions presented here can be found elsewhere [26].

The application of the Poisson equation to the cylindrical undoped device shown in Fig. 10.10 yields

$$\frac{d^2\phi}{dr^2} + \frac{1}{r} \frac{d\phi}{dr} = \frac{\phi_t}{L_{Di}^2} e^{\frac{\phi-V}{\phi_t}} \quad (10.4.1)$$

where V is the electron quasi-Fermi potential and

$$L_{Di} = \left(\frac{kT\epsilon_s}{q^2 n_i} \right)^{1/2}$$

is the intrinsic Debye length. In expression (10.4.1), the gradual-channel approximation is assumed to be valid and the hole density is considered negligible compared with the electron density.

Expression (10.4.1) must satisfy the following symmetric boundary condition

$$\frac{d\phi}{dr}(r=0) = 0. \quad (10.4.2)$$

The surface potential ϕ_s is

$$\phi_s = \phi(r=R). \quad (10.4.3)$$

Assuming that the current flows along the y-direction only, then V does not change along the r-direction. The solution of (10.4.1) with the symmetric boundary condition in (10.4.2) yields [25], [26], [27]

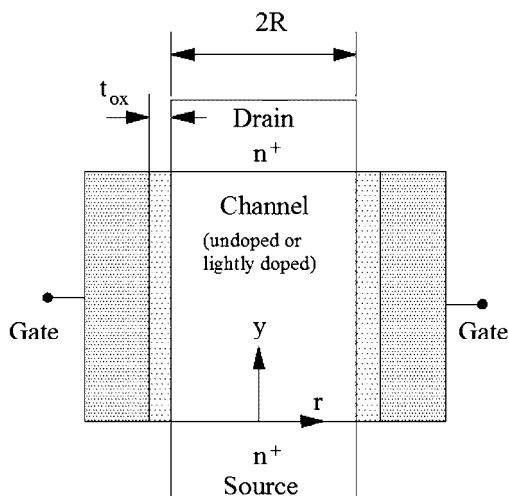


Fig. 10.10 Cross-section of a surrounding-gate MOSFET. (After [26].)

$$\phi(r) = V + \phi_t \ln \left(\frac{-8BL_{Di}^2}{(1 + Br^2)^2} \right) \quad (10.4.4)$$

where B is a constant that must be determined from the boundary conditions.

The expressions that follow are similar to those derived in [26] except for a slight change in the symbols. In order to find B in (10.4.4) we write the total electron charge per unit gate area as³

$$Q = -C_{ox} (V_G - \Delta\phi - \phi_s) \quad (10.4.5)$$

where $\Delta\phi$ is the work-function difference between the gate electrode and intrinsic silicon and $C_{ox} = \epsilon_{ox} / (R \ln(1 + t_{ox}/R))$. From Gauss' law we can write

$$Q = -\epsilon_s \left. \frac{d\phi}{dr} \right|_{r=R}. \quad (10.4.6)$$

³ As for SOI transistors we use $Q(C)$ instead of $Q'(C')$ for charge (capacitance) per unit area.

Equation (10.4.6), along with the derivative of the electrostatic potential calculated from (10.4.4), allows us to write

$$\left. \frac{d\phi}{dr} \right|_{r=R} = \frac{-4BR}{1+BR^2} \phi_t = -\frac{Q}{\epsilon_s}. \quad (10.4.7)$$

Expression (10.4.7) allows us to determine B in terms of the electron charge density Q enclosed in the cylinder. Now, inserting the value of B into (10.4.4) and recalling that $n = n_i \exp((\phi - V)/\phi_t)$ results in the following profile of electron concentration

$$n = n_i \frac{8 \frac{Q}{Q_0} \left(1 + \frac{Q}{Q_0} \right)}{\left(1 + \frac{Q}{Q_0} \left(1 - \frac{r^2}{R^2} \right) \right)^2} \frac{L_{Di}^2}{R^2} \quad (10.4.8)$$

where $Q_0 = -4\epsilon_s \phi_t / R$ is a normalization charge density. The electron concentration distribution depends strongly on the ratio Q/Q_0 , which is a measure of how inverted the device is. Deep in weak inversion, $Q/Q_0 \ll 1$, and both the electrostatic potential and the electron concentration are independent of the radius r , a characteristic of the volume-inversion phenomenon [28] that is not captured by standard charge-sheet models [29]. For increasing inversion levels, the electronic charge tends to be more concentrated at the silicon film periphery.

Now, inserting

$$\phi_s = \phi(r = R) = V + \phi_t \ln \left(\frac{-8BR^2}{(1+BR^2)^2} \frac{L_{Di}^2}{R^2} \right) \quad (10.4.9)$$

into (10.4.5) and using (10.4.7) we finally find that

$$\frac{V_G - \Delta\phi - V - 2\phi_t \ln(2L_{Di}/R)}{\phi_t} = \frac{\eta Q}{Q_0} + \ln \left[\frac{Q}{Q_0} \left(\frac{Q}{Q_0} + 1 \right) \right] \quad (10.4.10)$$

where $\eta = (4\epsilon_s / R) / C_{ox}$ is a structural parameter. Expression (10.4.10) represents a charge control model for the surrounding-gate MOSFET,

which is very similar to the UCCM previously derived for bulk and FD SOI transistors [26].

In weak inversion ($Q/Q_0 \ll 1$), the relationship between charge and voltage in (10.4.10) becomes

$$Q \cong -\frac{4\epsilon_s}{R} \phi_t \exp \left(\frac{V_G - \Delta\phi - V - 2\phi_t \ln(2\sqrt{2} L_{Di}/R)}{\phi_t} \right) \quad (10.4.11)$$

while, in strong inversion ($Q/Q_0 \gg 1$)

$$Q \cong -C_{ox} \left(V_G - \Delta\phi - V - 2\phi_t \ln(2\sqrt{2} L_{Di}/R) \right). \quad (10.4.12)$$

For cases where the accurate meaning of the equilibrium threshold voltage V_{T0} is not essential, we can define it as the gate voltage, for $V=0$, such that the extrapolated charge in (10.4.12) equals zero

$$V_{T0} = \Delta\phi + 2\phi_t \ln(2\sqrt{2} L_{Di}/R). \quad (10.4.13)$$

The drain current is calculated from the Pao-Sah expression in cylindrical coordinates

$$I_D = \mu_n \frac{2\pi R}{L} \int_{V_s}^{V_D} Q(V) dV. \quad (10.4.14)$$

The derivative of the voltage wrt to the charge derived from (10.4.10) is

$$\frac{-dV}{\phi_t} = \left[\frac{\eta}{Q_0} + \frac{1}{Q} + \frac{1}{Q+Q_0} \right] dQ, \quad (10.4.15)$$

which leads to

$$I_D = \mu_n \frac{2\pi R}{L} Q_0 \phi_t \left[\frac{\eta}{2} \frac{Q_s^2 - Q_D^2}{Q_0^2} + 2 \frac{Q_s - Q_D}{Q_0} - \ln \left(\frac{Q_s + Q_0}{Q_D + Q_0} \right) \right] \quad (10.4.16)$$

after substitution of (10.4.15) into (10.4.14).

The first term in the right-hand side of (10.4.16) is dominant in strong inversion; thus, the drain current is approximated by

$$I_D = \mu_n \frac{\pi R}{L} C_{ox} \left[(V_G - V_{T0} - V_s)^2 - (V_G - V_{T0} - V_D)^2 \right] \quad (10.4.17)$$

in strong inversion. On the other hand, both the second and third term of (10.4.16) must be accounted for in the subthreshold regime, yielding

$$I_D = \mu_n \frac{8\pi\epsilon_s}{L} \phi_t^2 \exp\left(\frac{V_G - V_{T0} - V_S}{\phi_t}\right) \left[1 - \exp\left(\frac{-V_{DS}}{\phi_t}\right)\right] \quad (10.4.18)$$

$$= \mu_n \frac{\pi R^2}{L} n_i kT \exp\left(\frac{V_G - \Delta\phi - V_S}{\phi_t}\right) \left[1 - \exp\left(\frac{-V_{DS}}{\phi_t}\right)\right].$$

In weak inversion, the current is proportional to the device area, owing to the uniform distribution of carriers along the cross-sectional area, and independent of the oxide thickness [26]. This is a characteristic of volume-inversion devices, where the carriers are distributed along the silicon film rather than concentrated at the film periphery.

10.5 Multiple-gate transistors

The International Technology Roadmap for Semiconductors anticipates that advanced non-classical MOSFETs, such as ultra-thin body multiple-gate structures with lightly doped channels, will be utilized for very short channels. The main goal of such devices is to reduce short-channel effects in deeply scaled technologies. Essentially, multiple-gate transistors reduce SCE by preventing the electric field lines originating at the drain from terminating under the channel [30].

Figure 10.11 shows some of the existing gate topologies for thin-film MOSFETs: 1) single gate; 2) double gate; 3) triple gate; 4) quadruple gate (or gate all-around); 5) Pi-gate (triple-gate) MOSFET [30].

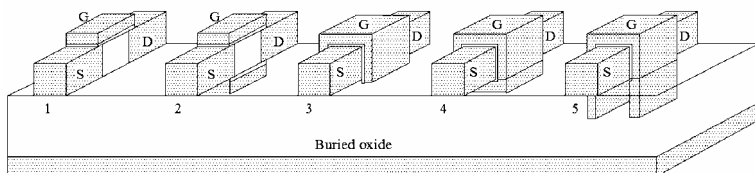


Fig. 10.11 Different gate configurations for multiple-gate devices: 1) single gate; 2) double gate; 3) triple gate; 4) quadruple gate, 5) Pi-gate MOSFET. (After [30].)

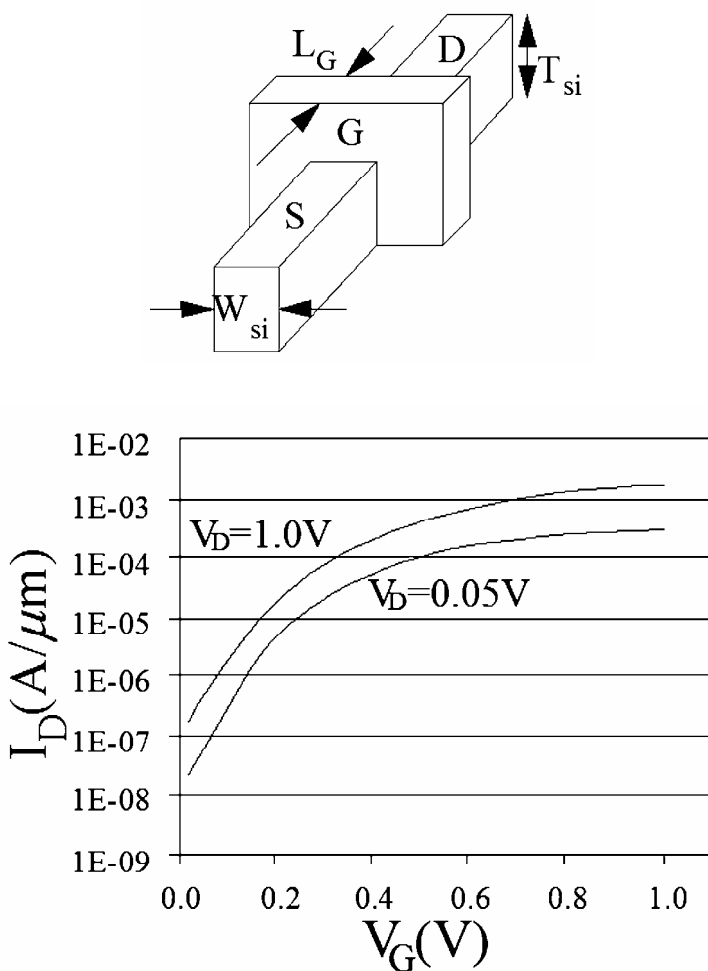


Fig. 10.12 Triple-gate non-planar transistor and current versus gate voltage characteristic of n-channel transistor: $L_G = 60$ nm, $T_{Si} = 36$ nm, $W_{Si} = 55$ nm, $T_{ox} = 1.5$ nm. (After [35], [36].)

Much research work has been published [30]-[39] on how to optimize the device performance by changing gate length, channel width, and silicon film thickness, besides channel doping in the case of SOI transistors. One of the most promising devices to be used in future CMOS technologies is the non-planar tri-gate transistor [35]-[38], whose

topology is sketched in Fig. 10.12 together with the current versus voltage characteristics of an n-channel device [35], [36] which has exhibited excellent performance as demonstrated by the close to ideal slope factor, DIBL, and large I_{on}/I_{off} ratio. The PMOS transistor has a subthreshold slope of 69.5 mV/decade, DIBL = 48 mV/V, I_{on} = 520 $\mu\text{A}/\mu\text{m}$ and I_{off} = 24 nA/ μm at 1.3 V.

Bibliography

- [1] International Technology Roadmap for Semiconductors (2004 update). [Online]. Available: <http://public.itrs.net/>
- [2] H.-S. P. Wong, "Beyond the conventional transistor," *IBM J. Res. Dev.*, vol. 46, no. 2/3, pp. 133-168, Mar./May 2002.
- [3] Yuan Taur, Douglas A. Buchanan, Wei Chen, David J. Frank, Khalid E. Ismail, Shih-Hsien Lo, George A. Sai-Halasz, Raman G. Viswanathan, Hsing-Jen C. Wann, Shalom J. Wind, Hon-Sum Wong, "CMOS scaling into the nanometer regime," *Proc. IEEE*, vol. 85, no. 4, pp. 486-504, Apr. 1997.
- [4] Yuan Taur, David J. Frank, H.-S. Philip Wong, Douglas A. Buchanan, Shalom J. Wind, Stephen A. Rishton, George A. Sai-Halasz, Edward J. Nowak, "CMOS scaling into the 21st century: 0.1 μm and beyond," *IBM J. Res. Dev.*, vol. 39, no. 1/2, pp. 245-260, Jan./Mar. 1995.
- [5] Rafael Rios, Wei-Kai Shih, Atul Shah, Sivakumar Mudanai, Paul Packan, Tracy Sandford, and Kaizard Mistry, "A three-transistor threshold voltage model for halo processes," *IEDM Tech. Dig.*, 2002, pp. 113-116.
- [6] Takashi Hori, "A 0.1 μm CMOS technology with Tilt-Implanted Punchthrough Stopper (TIPS)," *IEDM Tech. Dig.*, 1994, pp. 75-78.
- [7] Yuan Taur and Edward J. Nowak, "CMOS devices below 0.1 μm : How high will performance go?," *IEDM Tech. Dig.*, 1997, pp. 215-218.
- [8] Hioraki Ueno, Daisuke Kitamaru, Keiichi Morikawa, Masayasu Tanaka, Mitiko Miura-Mattausch, Hans Juergen Mattausch, Shigetaka Kumashiro, Tetsuya Yamaguchi, Kyoji Yamashita, and Noriaki Nakayama, "Impurity-profile-based threshold-voltage model of pocket-implanted MOSFETs for circuit simulation," *IEEE Trans. Electron Devices*, vol. 49, no. 10, pp. 1783-1789, Oct. 2002.
- [9] Y. Okumura, M. Shirahata, T. Okudaira, A. Hachisuka, H. Arima, T. Matsukawa, and N. Tsubouchi, "A novel source-to-drain nonuniformly doped channel (NUDC) MOSFET for high-current drivability and threshold voltage controllability," *IEDM Tech. Dig.*, 1990, pp. 391-394.
- [10] G. G. Shahidi, "SOI technology for the GHz era," *IBM J. Res. Dev.*, vol. 46, no. 2/3, pp. 121-131, Mar./May 2002.
- [11] K. Hashimoto, T. I. Kamins, K. M. Cham, and S. Y. Chiang, "Characteristics of sub-micron CMOS transistors in implanted-buried-oxide SOI films," *IEDM Tech. Dig.*, 1985, pp. 672-675.

- [12] Yuan Taur and Tak H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.
- [13] Bernard M. Tenbroek, Michael S. L. Lee, William Redman-White, R. John T. Bunyan, and Michael J. Uren, "Impact of self-heating and thermal coupling on analog circuits in SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 33, no. 7, pp. 1037-1046, July 1998.
- [14] Lisa T. Su, Jarvis B. Jacobs, James E. Chung, and Dimitri A. Antoniadis, "Deep-submicrometer channel design in silicon-on-insulator (SOI) MOSFET's," *IEEE Electron Device Lett.*, vol. 15, no. 9, pp. 366-369, Sep. 1994.
- [15] Hyung-Kyu Lim and Jerry G. Fossum, "Current-voltage characteristics of thin-film SOI MOSFET's in strong inversion," *IEEE Trans. Electron Devices*, vol. 31, no. 4, pp. 401-408, Apr. 1984.
- [16] Hyung-Kyu Lim and Jerry G. Fossum, "A charge-based large-signal model for thin-film SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 32, no. 2, pp. 446-457, Feb. 1985.
- [17] Adelmo Ortiz-Conde, Francisco J. García Sánchez, Pierre E. Schmidt, and Augusto Sa-Neto, "The nonequilibrium inversion layer charge of the thin-film SOI MOSFET," *IEEE Trans. Electron Devices*, vol. 36, no. 9, pp. 1651-1656, Sep. 1989.
- [18] C. Mallikarjun and K. N. Bhat, "Numerical and charge sheet models for thin-film SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 37, no. 9, pp. 2039-2051, Sep. 1990.
- [19] Benjamín Iníiguez, Luiz Fernando Ferreira, Bernard Gentinne, and Denis Flandre, "A physically-based C_{∞} -continuous fully-depleted SOI MOSFET model for analog applications," *IEEE Trans. Electron Devices*, vol. 43, no. 4, pp. 568-575, Apr. 1996.
- [20] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510-1519, Oct. 1998.
- [21] A. Ortiz-Conde, F. J. García Sánchez, P. E. Schmidt and A. Sa-Neto, "The foundation of a charge-sheet model for the thin-film MOSFET," *Solid-State Electron.*, vol. 31, no. 10, pp. 1497-1500, Oct. 1988.
- [22] Hiroshi Takato, Kazumasa Sunouchi, Naoko Okabe, Akihiro Nitayama, Katsuhiko Hieda, Fumio Horiguchi, and Fujio Masuoka, "Impact of surrounding gate transistor (SGT) for ultra-high-density LSI's," *IEEE Trans. Electron Devices*, vol. 38, no. 3, pp. 573-578, Mar. 1991.
- [23] Shigeyoshi Watanabe, Kenji Tsuchida, Daisaburo Takashima, Yukihiro Oowaki, Akihiro Nitayama, Katsuhiko Hieda, Hirishi Takato, Kazumasa Sunouchi, Fumio Horiguchi, Kazunori Ohuchi, Fujio Masuoka, and Hisashi Hara, "A novel circuit technology with surrounding gate transistors (SGT's) for ultra- high density DRAM's," *IEEE J. Solid-State Circuits*, vol. 30, no. 9, pp. 960-971, Sep. 1995.
- [24] Christopher P. Auth and James D. Plummer, "Scaling theory for cylindrical fully-depleted surrounding- gate MOSFET's," *IEEE Electron Device Lett.*, vol. 18, no. 2, pp. 74-76, Feb. 1997.
- [25] Yijian Chen and Jianfeng. Luo, "A comparative study of double-gate and surrounding-gate MOSFETs in strong inversion and accumulation using an

- analytical model,” in *Proc. Int. Conf. on Modeling and Simulation of Microsystems*, 2001, pp. 546–549.
- [26] Benjamin Iñíguez, David Jiménez, Jaume Roig, Hamdy A. Hamid, Lluís F. Marsal, Josep Pallarès, “Explicit continuous model for long-channel undoped surrounding gate MOSFETs,” *IEEE Trans. Electron Devices*, vol. 52, no. 8, pp. 1868–1873, Aug. 2005.
- [27] P. L. Chambré, “On the solution of the Poisson-Boltzmann equation with application to the theory of thermal explosions,” *J. Chem. Phys.*, Vol. 20, no. 11, pp. 1795–1797, Nov. 1952.
- [28] J. Brini, M. Benachir, G. Ghibaudo, F. Balestra, “Threshold voltage and subthreshold slope of the volume-inversion MOS transistor,” *IEE Proc.-G*, vol. 138, no. 1, pp. 133–136, Feb. 1991.
- [29] D. Jiménez, B. Iñíguez, J. Suñé, L. F. Marsal, J. Pallarès, J. Roig, and D. Flores, “Continuous analytic I-V model for surrounding-gate MOSFETs,” *IEEE Electron Device Lett.*, vol. 25, no. 8, pp. 571–573, Aug. 2004.
- [30] Jong-Tae Park and Jean-Pierre Colinge, “Multiple-gate SOI MOSFETs: device design guidelines,” *IEEE Trans. Electron Devices*, vol. 49, no. 12, pp. 2222–2229, Dec. 2002.
- [31] Qiang Chen, Evans M. Harrell II, and James D. Meindl, “A physical short-channel threshold voltage model for undoped symmetric double-gate MOSFETs,” *IEEE Trans. Electron Devices*, vol. 50, no. 7, pp. 1631–1637, July 2003.
- [32] Yuan Taur, “An analytical solution to a double-gate MOSFET with undoped body,” *IEEE Electron Device Lett.*, vol. 21, no. 5, pp. 245–247, May 2000.
- [33] Yuan Taur, Xiaoping Liang, Wei Wang, and Huaxin Lu, “A continuous analytic drain-current model for DG MOSFETs,” *IEEE Electron Device Lett.*, vol. 25, no. 2, pp. 107–109, Feb. 2004.
- [34] Xiaoping Liang and Yuan Taur, “A 2-D analytical solution for SCEs in DG MOSFETs,” *IEEE Trans. Electron Devices*, vol. 51, no. 9, pp. 1385–1391, Sep. 2004.
- [35] B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, “High performance fully-depleted tri-gate CMOS transistors,” *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 263–265, Apr. 2003.
- [36] B. Doyle, B. Boyanov, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, R. Rios and R. Chau, “Tri-gate fully-depleted CMOS transistors: fabrication, design and layout,” *Symp. VLSI Tech.*, 2003, pp. 133–134.
- [37] Robert Chau, Boyan Boyanov, Brian Doyle, Mark Doczy, Suman Datta, Scott Hareland, Ben Jin, Jack Kavalieros, and Matthew Metz, “Silicon nano-transistors for logic applications,” *Physica E*, vol. 19, no. 1–2, pp. 1–5, July 2003.
- [38] Brian Doyle, Reza Arghavani, Doug Barlage, Suman Datta, Mark Doczy, Jack Kavalieros, Anand Murthy, and Robert Chau, “Transistor elements for 30nm physical gate lengths and beyond,” *Intel Tech. J.*, vol. 6, no.2, pp. 42–54, May 2002.
- [39] Xuejue Huang, Wen-Chin Lee, Charles Kuo, Digh Hisamoto, Leland Chang, Jakub Kedzierski, Erik Anderson, Hideki Takeuchi, Yang-Kyu Choi, Kazuya Asano, Vivek Subramanian, Tsu-Jae King, Jeffrey Bokor, and Chenming Hu, “Sub 50-nm FinFET: PMOS,” *IEDM Tech. Dig.*, 1999, p. 67–70.

Problems

10.1. Estimate the threshold voltage of a fully-depleted n-channel SOI transistor with the following parameters: $t_{ox}=5$ nm, $t_{Si}=30$ nm, $N_A=2 \times 10^{18}$ cm³, $V_{FB}=-0.9$ V. Assume the back oxide capacitance to be negligible.

10.2. Assume that an undoped surrounding gate transistor has the following parameters: $R=40$ nm, $L_{Di}=40$ μ m. Assume also that $V=0$. Plot the curves $\phi(r)/\phi_i$ vs r/R and $n(r)/n_i$ vs r/R for $Q/Q_0=0.1, 1, 10$, and 100 . As you can see, as the inversion charge density decreases, the carriers tend to occupy the whole volume of the cylinder rather than concentrate at the surface.

Chapter 11

MOSFET Parameter Extraction

MOSFET parameters are fundamental for the correct analysis, design, and simulation of MOS circuits. The accuracy of the transistor characteristics depends not only on a good device model but also on the accuracy of fundamental parameters. This chapter describes some procedures to extract fundamental parameters of MOSFET models.

11.1 Introduction

The consistent and accurate determination of model parameters plays an important role in device design, process control, and technology characterization [1]. Also, the transistor model and the associated set of model parameters are of utmost relevance for interfacing integrated circuit designers and silicon foundries. The accuracy of the device characteristics and, as a consequence, the prediction of the performance of a circuit depends not only on the device model but also on the parameter values being used [2], [3]. Hence, the procedures employed to extract the device model parameters are of prime importance.

Determining the values for the full set of parameters of a MOSFET model is not a simple task. As pointed out in [2] and [3], difficulties in determining the values for the model parameters arise because: (i) the approximations employed to derive the device model are, in some cases, far from reality; (ii) the accurate determination of a given parameter sometimes depends on the value of another parameter or parameters which is or are not accurately known. The extraction of a full set of transistor parameters, a must in a production environment, generally involves some kind of optimization, which is, however, beyond the scope

of this text. The extraction of parameters presented here, limited to a set of dc parameters of the intrinsic transistor, is based on a guiding principle enunciated in [3] : a common way to determine a parameter is to concentrate on a particular operating region where this parameter has a dominant effect. This guiding principle avoids the need for the accurate determination of other parameters, gives an insight into how that specific parameter affects the transistor characteristics, and generally results in an extraction procedure that is simple.

11.2 Threshold voltage and short-channel effects

The threshold voltage V_T is a fundamental parameter in the modeling and characterization of MOS transistors. Different definitions of threshold voltage have been presented in the open literature [4]. Sometimes, qualitative concepts such as “ V_T represents the onset of conduction in the channel or the onset of significant drain current flow” are given as definitions. Even if these definitions are appealing, they have limited value because of their ambiguity. To make things worse, most of the studies on extraction use a simple model of the drain current limited to a specific operating region. To analyze a V_T extraction procedure, it is essential to use a model that includes both the drift and diffusion transport mechanisms, because both phenomena are important near the threshold condition.

At least a dozen methods [1], [4], [5], [6], to extract the threshold voltage, such as the traditional ELR (Extrapolation in the Linear Region), the TC/SDL (Transconductance Change/Second Derivative Logarithmic), the constant-current (CC), and the g_m/I_D (transconductance-to-current ratio), are available. The ELR, although simple, is not accurate since it is based on the extrapolated current versus gate voltage curve in strong inversion. The extrapolated threshold voltage determined using the ELR method is ambiguous since the drain current versus gate voltage characteristic is not a straight line; thus, the tangent to the curve depends on the choice of the operating point (see Fig. 11.1).

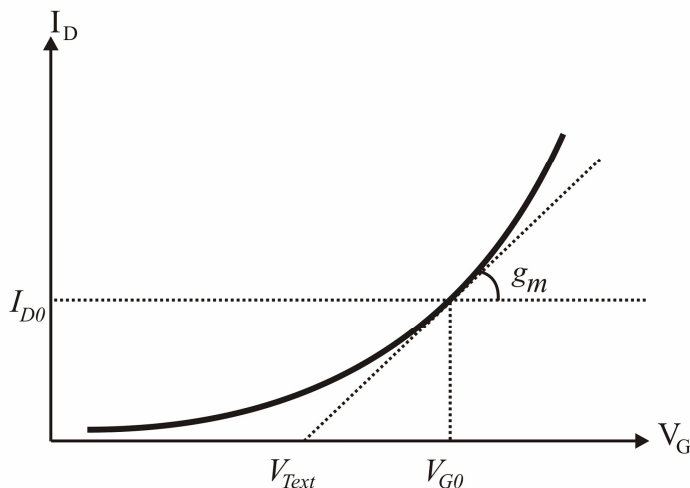


Fig. 11.1 Determination of the extrapolated threshold voltage (V_{Text}) using the ELR method.

Contrary to the ELR method, the TC, SDL, CC, and g_m/I_D methods have the advantage of being based on a physical property of the device, thus being independent of a specific model, and they determine univocally the threshold voltage. Unfortunately, both the TC and SDL methods are very sensitive to noise in the measurements and in the numerical processing. On the other hand, the simplest method, the CC method, relies on the definition of the specific current per aspect ratio and on the proper determination of the channel length and width.

In the following, a methodology for the extraction of the threshold voltage [7] based on the MOSFET model developed in previous chapters is presented. The procedure proposed for the extraction of V_T , which is based on the transconductance-to-current ratio, is performed over the weak and moderate inversion regions with small drain-source voltages so that short channel effects impose no drawbacks. Even series resistances of source and drain may be disregarded owing to the negligible voltage drops across them. The threshold voltage is determined according to a clear physical definition and its value closely agrees with the threshold voltage extracted through other methodologies [7].

The procedure described next satisfies the three major criteria for the definition and extraction of V_T , namely simplicity, unambiguity, and consistency [5].

Neglecting the dependence of I_S on gate voltage, the transconductance-to-current ratio, Eq. (3.7.27), becomes

$$\frac{g_{mg}}{I_D} = \frac{1}{i_f - i_r} \left(\frac{\partial i_f}{\partial V_G} - \frac{\partial i_r}{\partial V_G} \right) = \frac{2}{n\phi_t (\sqrt{1+i_f} + \sqrt{1+i_r})}. \quad (11.2.1)$$

For V_{DS} much lower than the thermal voltage ϕ_t , $i_f \cong i_r$ and (11.2.1) becomes

$$\frac{g_{mg}}{I_D} \cong \frac{1}{n\phi_t \sqrt{1+i_f}}. \quad (11.2.2)$$

Now, if one assumes n to be almost constant over the measurement range, (11.2.2) can be written as

$$\frac{g_{mg}}{I_D} \cong \frac{g_{mg}}{I_D} \bigg|_{\max} \frac{1}{\sqrt{1+i_f}}. \quad (11.2.3)$$

In expression (11.2.3), the maximum transconductance-to-current ratio is the value of the transconductance-to-current ratio deep in weak inversion ($i_f \rightarrow 0$). For $i_f = 3$, the transconductance-to-current ratio is one-half of its peak value.

Equation (11.2.3) is the basis of a very simple and quick method for determining both the threshold voltage and the specific current I_S using a single current-voltage characteristic, as will be shown next. For a MOSFET biased at low V_{DS} , (11.2.3) shows that the deviation of the transconductance-to-current ratio from the maximum value in weak inversion depends on the inversion level only. The slight variations of the slope factor and mobility with gate voltage are negligible over the measurement range.

For both long-channel and short-channel transistors submitted to low V_{DS} voltages we rewrite below Eq. (3.7.20) for the sake of convenience

$$\frac{V_{DS}}{\phi_t} = \sqrt{1+i_f} - \sqrt{1+i_r} + \ln \frac{\sqrt{1+i_f} - 1}{\sqrt{1+i_r} - 1}. \quad (11.2.4)$$

For $i_r \rightarrow i_f$ expression (11.2.4) becomes

$$\frac{V_{DS}}{\phi_t} \approx \frac{1}{2} \frac{i_f - i_r}{\sqrt{1+i_f} - 1} = \frac{1}{2} \frac{I_D / I_S}{\sqrt{1+i_f} - 1}. \quad (11.2.5)$$

The measurement of both V_T and I_S is taken from the g_{mgs}/I_D vs. V_G characteristic according to expressions (11.2.3) and (11.2.5). The gate voltage at which the value of g_{mgs}/I_D drops to one-half of its peak value corresponds to $i_f = 3$. Recalling that for $i_f = 3$ and $V_S=0$, the pinch-off voltage $V_P=0$ or, equivalently, $V_G=V_{T0}$, the equilibrium threshold voltage V_{T0} is the value of V_G at which $g_{mgs}/I_D \approx 0.5 (g_{mgs}/I_D)_{max}$. Now, using (11.2.5) we find that the value of the specific current is, for $i_f = 3$, calculated as

$$I_S \approx \frac{\phi_t}{2V_{DS}} \frac{I_D}{\sqrt{1+i_f} - 1} \bigg|_{i_f=3} = \frac{\phi_t}{2V_{DS}} I_D. \quad (11.2.6)$$

If one chooses $V_{DS}/\phi_t=1/2$, then $I_S=I_D$. The circuit for the determination of both the threshold voltage and the specific current is shown in Fig. 11.2. Figure 11.3 displays the g_{mgs}/I_D vs. V_G characteristic. The circle in Fig. 11.3 corresponds to the point at which g_{mgs}/I_D drops to one-half of its peak value. The gate voltage thus measured is the equilibrium threshold voltage and the corresponding current is the specific current I_S .

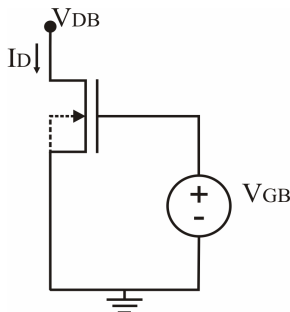


Fig. 11.2 Circuit configuration for measuring the common-source characteristics.

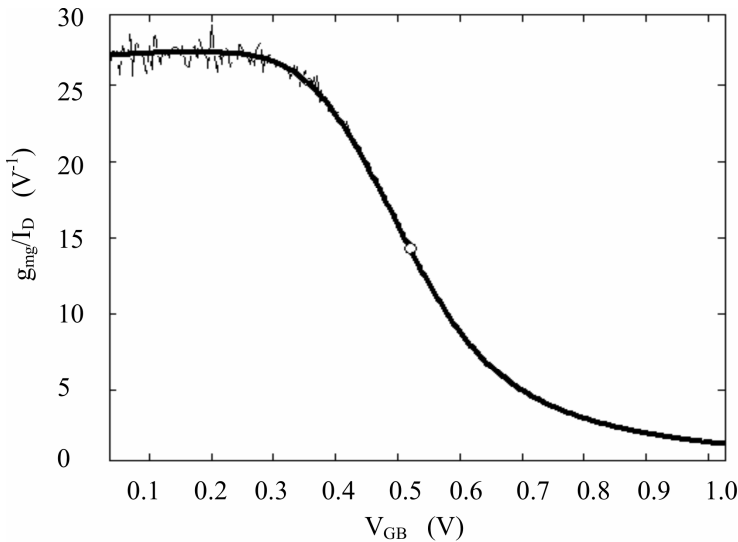


Fig. 11.3 Transconductance-to-current ratio for $V_{DS} \cong \phi/2$ and $V_S=0$. Dotted line: measured g_{mg}/I_D ; solid line: filtered g_{mg}/I_D ; circle: $g_{mg}/I_D \cong 0.5 (g_{mg}/I_D)_{max}$. $L_m = 0.2 \mu m$ (mask channel length), $W_m = 20 \mu m$ (mask channel width) in $0.18 \mu m$ CMOS technology.

In order to account for the error introduced by a non-negligible difference between the forward and reverse currents, a small correction can be included in expressions (11.2.3) and (11.2.6), as shown in [7].

For modeling purposes, the determination of the reduction of threshold voltage with decreasing channel length, named short-channel effect (SCE), and with increasing drain voltage, called DIBL, is also required. The threshold voltage shift due to SCE measured at a given drain-to-source voltage is defined [9] as

$$\Delta V_T(L) = V_T(L) - V_T(L \rightarrow \infty), \quad (11.2.7)$$

where L is the channel length. The threshold shift due to DIBL for a given channel-length is given by

$$\Delta V_T(DIBL) = V_G(V_{DS} = V_{DD})|_{I_D} - V_G(V_{DS} = 50mV)|_{I_D}. \quad (11.2.8)$$

In (11.2.8), the gate voltages are measured for a specified drain current, usually a value around the specific current I_S , which means that the measured values of gate voltages are around the threshold voltage.

V_{DD} is the maximum voltage for the technology under evaluation. Fig. 11.4 illustrates the dependence of the threshold voltage on the channel length for pocket-implanted and non pocket-implanted devices for two values of drain-to-source voltage. The threshold shifts can be readily determined from the curves shown.

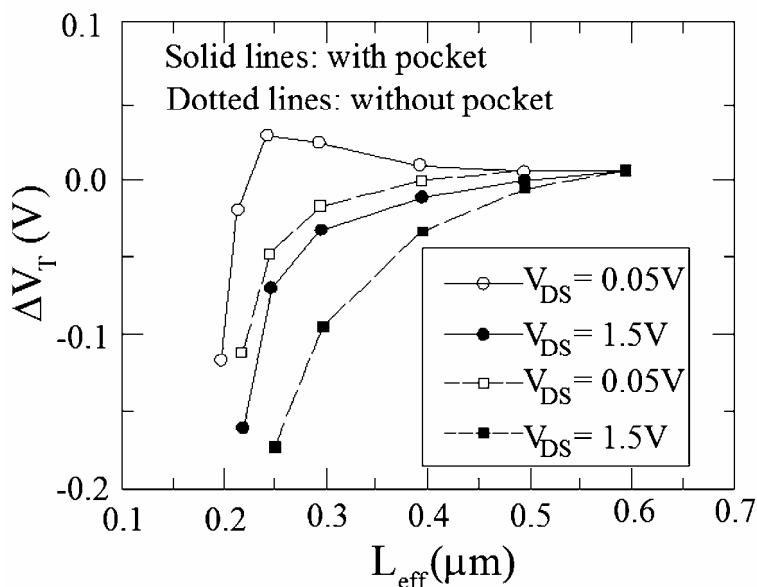


Fig. 11.4 Threshold voltage variation as a function of channel length and V_{DS} . (After [9].)

11.3 Specific current and effective channel length and width

We have demonstrated that, for a uniformly doped substrate, the specific current is proportional to the aspect ratio W/L . If the methodology for extracting I_S is performed for several large-width devices with different channel lengths, the plot of the reciprocal of the measured specific current per unit width vs. mask channel length, *i.e.*, W/I_S vs. L_m (circles in Fig. 11.5), allows the determination of the channel length shift $\Delta L = L_m - L$, where L_m is the mask channel length and L is the electrical channel length [8].

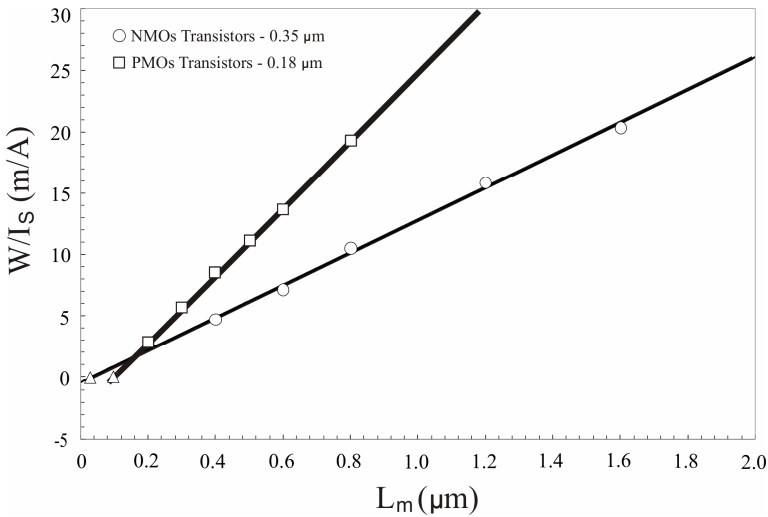


Fig. 11.5 Reciprocal of the specific current per unit width vs. mask channel length. Circles and squares are extracted values of W/I_S ; solid line: fitting straight-line; triangle: $L_m = \Delta L$. NMOS and PMOS devices were fabricated in 0.35 μm and 0.18 μm CMOS technologies [8], respectively.

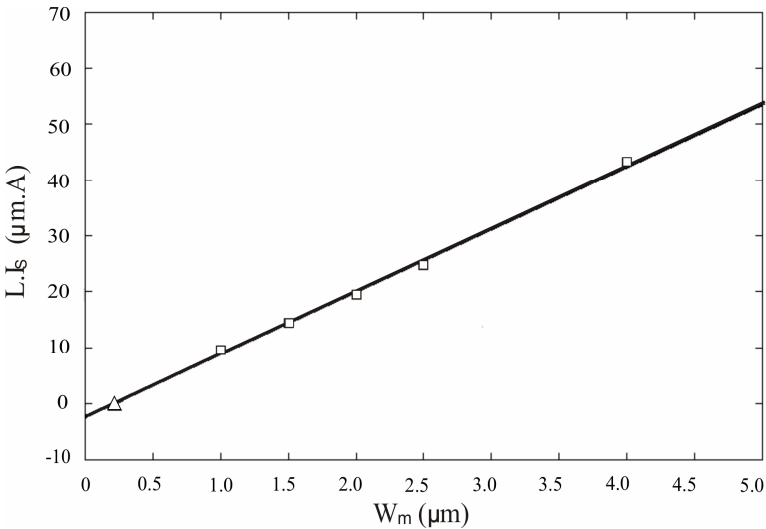


Fig. 11.6 Specific current length product vs. mask channel width. Squares: extracted values of $L \cdot I_S$; solid line: fitted straight-line; triangle: $W_m = \Delta W$. NMOS devices were fabricated in a 0.35 μm CMOS technology [8].

Similarly, the shift of the channel width $\Delta W = W_m - W$ may be evaluated through the fitting of a straight-line to the plot of $L \cdot I_S$ vs. W_m , where I_S is extracted for several devices with different channel widths, as shown in Fig. 11.6 [8].

11.4 Slope factor and subthreshold slope

One important parameter for the characterization of MOS transistors is the slope factor. The “ideal” slope factor is equal to one. Some of the structures we have presented in Chapter 10 have close to ideal slope factors; the bulk MOS transistor, however, is characterized by a slope factor a few percent to tens of percent higher than one. As previously explained, the reason for the deviation from the ideal slope factor in the bulk transistor is that a change in the gate voltage is not only accompanied by a change in the inversion charge but also by a change in the bulk charge.

Some methods can be devised to measure the slope factor. For long-channel devices, a very simple methodology is described in [10], [11] and repeated here. To explain the procedure employed to determine V_P and n , let us recall the UCCM:

$$\frac{V_P - V_S}{\phi_t} = \sqrt{1 + i_f} - 2 + \ln(\sqrt{1 + i_f} - 1). \quad (11.4.1)$$

Let us assume that the transistor specific current is constant; thus, the constant bias current in the common-source amplifier shown in Fig. 11.7 implies that $i_f = 3$. An examination of (11.4.1) shows that $V_P = V_S$ for $i_f = 3$. Hence, the value of the source voltage in the circuit of Fig. 11.7 gives the pinch-off voltage for the measured value of the gate voltage. The result of the measurement of the variation of the pinch-off voltage as well as $n = 1/(dV_P/dV_G)$ is shown in Fig. 11.7. Recalling that for a uniformly doped transistor the slope factor is

$$n \cong 1 + \frac{\gamma}{2\sqrt{V_P} + 2\phi_F}, \quad (11.4.2)$$

the variation of n with V_P can be used to extract both the body-effect factor and the Fermi level.

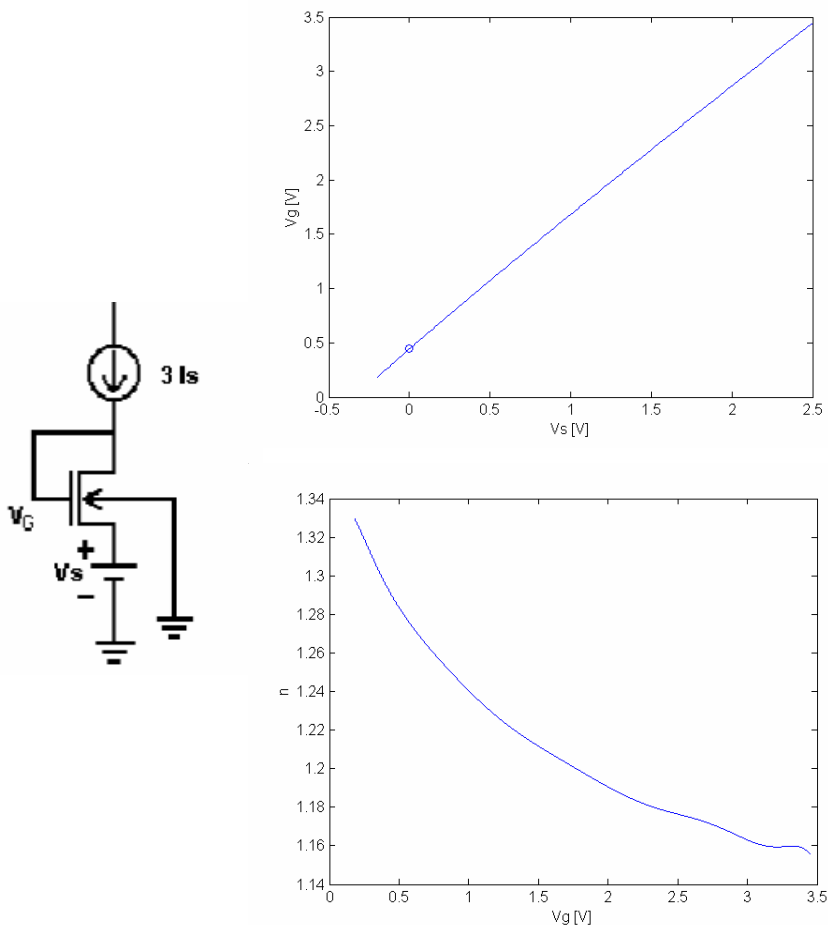


Fig. 11.7 Common-source circuit driven by a constant current source used for the determination of the pinch-off voltage and the slope factor, as functions of V_G . NMOS transistor $W=20\text{ }\mu\text{m}$, $L=2\text{ }\mu\text{m}$, $0.18\text{ }\mu\text{m}$ CMOS technology.

The subthreshold slope (SS), which gives the percent variation of the drain current as a function of the gate voltage in weak inversion, was extracted from measurements as shown in Fig. 11.8 for the same device described in Fig. 11. 7. $SS \cong 77\text{ mV/decade}$, which is equivalent to

$n \cong 1.3$ at 25°C . As can be noted, $n \cong 1.3$ is also the value of the slope factor for low values of gate voltage, *i.e.*, in weak inversion, that can be seen in Fig. 11.7.

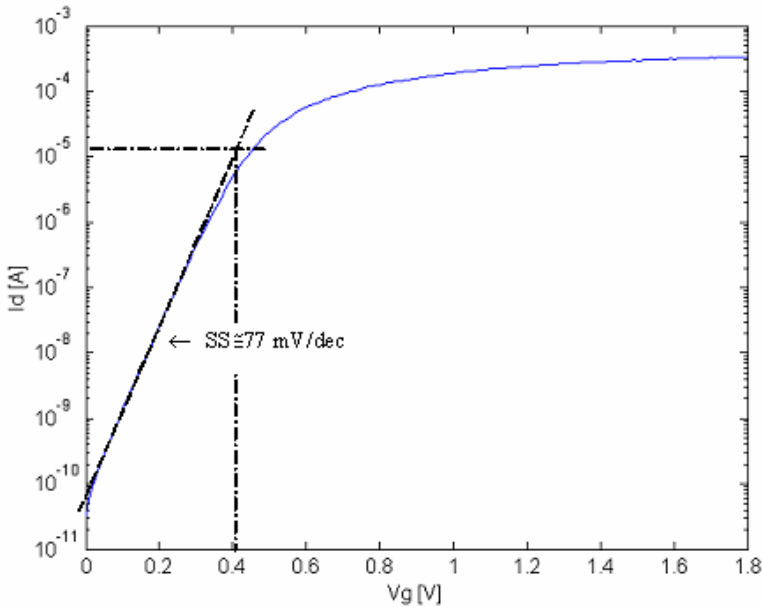


Fig. 11.8 Drain-current vs. gate voltage and determination of the subthreshold slope.

11.5 Mobility degradation with transversal field

In this section, we assume the MOSFET to be a long-channel device. Thus, the mobility degradation is caused by the transversal field only since the longitudinal field is assumed to be very small.

The Pao-Sah expression of the current is

$$I_D = \mu(y) W Q'_I \frac{dV}{dy}. \quad (11.5.1)$$

Since the mobility is no longer constant, the integration of (11.5.1) along the channel results in

$$I_D = \mu_{eff} \frac{W}{L} \int_{V_s}^{V_D} Q'_I(V) dV, \quad (11.5.2)$$

where

$$\mu_{eff} = \frac{L}{\int_0^L \frac{dy}{\mu(y)}}. \quad (11.5.3)$$

Now, assuming as in Chapter 4 [3], [12] the mobility to be given by

$$\mu(y) = \frac{\mu_0}{1 + \alpha_\theta F_{ave}(y)} \quad (11.5.4)$$

and the average electric field

$$F_{ave} = -\frac{Q'_B + \eta Q'_I}{\epsilon_s}, \quad (11.5.5)$$

the following expression results for the effective mobility

$$\mu_{eff} = \frac{\mu_0}{1 - \alpha_\theta \left(\frac{Q_B + \eta Q_I}{\epsilon_s WL} \right)}. \quad (11.5.6)$$

The term in parenthesis in the denominator of (11.5.6) is the average along the channel of the average fields in the x-direction. Q_B and Q_I are the total depletion and inversion charges in the semiconductor.

In order to determine parameter α_θ associated with mobility degradation, we carry out the following measurement procedures:

(i) Measure the variation of the normalization current I_S with respect to the gate voltage. The value of $I_S(V_G)$ can be determined from the circuit in Fig. 11.9 (a) [13]. To describe the methodology, we rewrite the expression of the source-transconductance-to-current ratio derived in Chapter 4

$$\frac{g_{ms} \phi_t}{I_F} = \frac{2}{\sqrt{1 + i_f} + 1}. \quad (11.5.7)$$

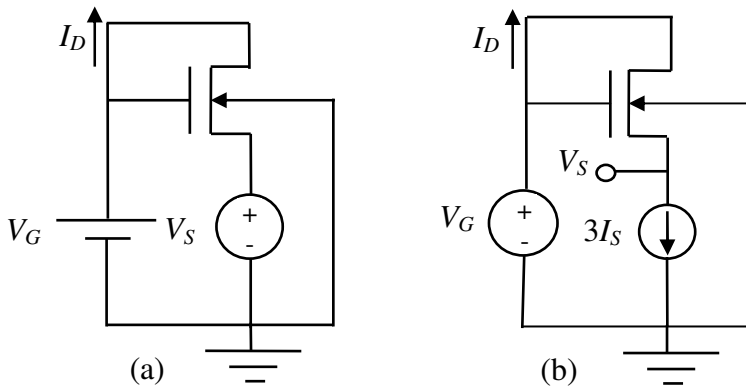


Fig. 11.9 Circuits for the measurement of (a) the specific current and (b) the slope factor.

Since the transistor in Fig. 11.9 operates in saturation, $I_F \cong I_D$. To measure I_S , we note that for $i_f=3$, the transconductance-to-current ratio drops to $2/3$ of its maximum value [13]. The specific current is $1/3$ of the drain current thus measured. Fig. 11.10 shows an example of the variation of the specific current with respect to the gate voltage. The specific current ranges from 22 nA to 35 nA for a gate voltage variation between 0.5 V and 5 V.

(ii) Measure the slope factor $n = 1/(dV_S / dV_G)$ using either the circuit in Fig. 11.7 or the one in Fig. 11.9 (b). Owing to the low sensitivity of the slope factor to the drain current, the accuracy of the drain current is not important.

(iii) Calculate the mobility from

$$\mu_{eff} = \frac{I_S}{n} \frac{1}{(W/2L) C'_{ox} \phi_t^2}. \quad (11.5.8)$$

Now, we write expression (11.5.6) as

$$\mu_{eff} \cong \frac{\mu_0}{1 - \alpha_\theta \left(\frac{Q_B}{\epsilon_s WL} \right)} \cong \frac{\mu_0}{1 + \theta \sqrt{V_P + 2\phi_F}}. \quad (11.5.9)$$

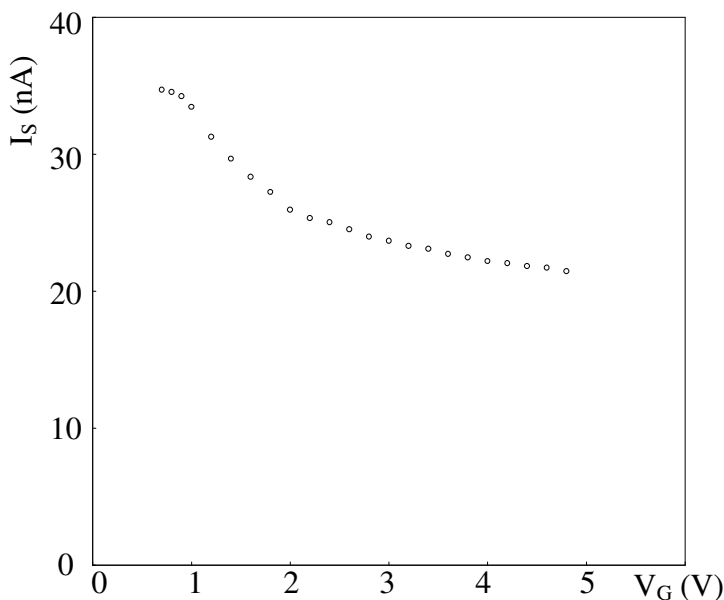


Fig. 11.10 Specific current of an NMOS transistor, 0.75 μm CMOS technology ($t_{\text{ox}} = 280 \text{ \AA}$), $W=L=25 \mu\text{m}$, versus gate voltage.

where

$$\theta = \frac{\alpha_{\theta} \gamma C'_{\text{ox}}}{\epsilon_s} \quad (11.5.10)$$

is the mobility degradation factor. To derive (11.5.9) we have assumed that the depletion charge is much larger than the inversion charge, a quite reasonable approximation for the low inversion level ($i_f=3$) for which we perform the measurements. In addition, we have approximated the depletion charge density in the channel by a constant value, which is also acceptable for low inversion levels. Finally, we have used the expression $Q'_B \cong -\gamma C'_{\text{ox}} \sqrt{V_p + 2\phi_F}$ derived in Chapter 3 for the depletion charge density.

The mobility calculated from measurements of specific current and slope factor is represented by open circles in Fig. 11.11, while the solid line is given by the fitting expression (11.5.9), with $\mu_0 = 645 \text{ cm}^2/\text{V}\cdot\text{s}$ and $\theta = 0.21 \text{ V}^{-1/2}$.

$$\mu C'_{ox} \frac{W}{L} \text{ (A/V}^2\text{)} \times 10^{-5}$$

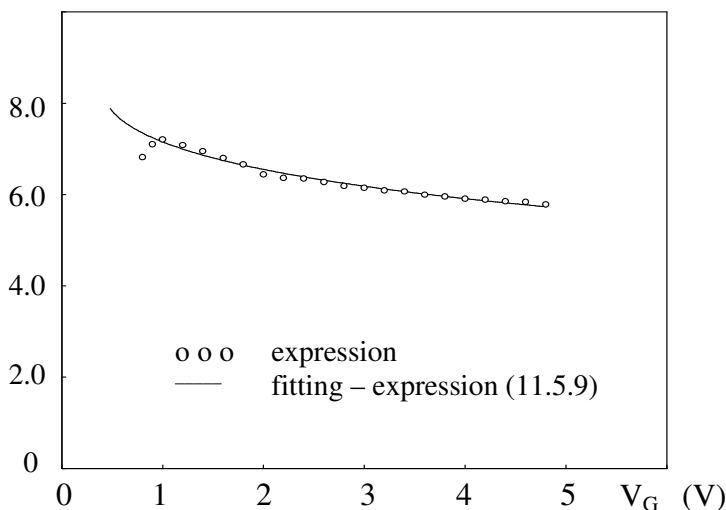


Fig. 11.11 Transconductance parameter ($\mu C'_{ox} W / L$) of an NMOS transistor, 0.75 μm CMOS technology ($t_{ox} = 280 \text{ \AA}$), $W=L=25 \mu\text{m}$, versus gate voltage, extracted from specific current and slope factor measurements.

Bibliography

- [1] Colin C. McAndrew and Paul A. Layman, "MOSFET effective channel length, threshold voltage, and series resistance determination by robust optimization," *IEEE Trans. Electron Devices*, vol. 39, no. 10, pp. 2298-2311, Oct. 1992.
- [2] Narain Arora, *MOSFET Models for VLSI Circuit Simulation Theory and Practice*, Springer-Verlag, Wien, 1993.
- [3] Yannis Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed., McGraw-Hill, Boston, 1999.
- [4] A. Ortiz-Conde, F. J. García Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectronics Reliability*, vol. 42, no. 4-5, pp. 583-596, Apr-May 2002.
- [5] X. Zhou, K. Y. Lim, and D. Lim, "A simple and unambiguous definition of threshold voltage and its implications in deep-submicron MOS device modeling," *IEEE Trans. Electron Devices*, vol. 46, no. 4, pp. 807-809, Apr. 1999.
- [6] Morikazu Tsuno, Masato Suga, Masayasu Tanaka, Kentaro Shibahara, Mitiko Miura-Mattausch, and Masataka Hirose, "Physically-based threshold voltage determination for MOSFET's of all gate lengths," *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1429-1434, July 1999.

- [7] A. I. A. Cunha, M. C. Schneider, C. Galup-Montoro, C. D. C. Caetano, and M. B. Machado, "Unambiguous extraction of threshold voltage based on the transconductance-to-current ratio," *Proc. Workshop on Compact Modeling, Nanotech 2005*, pp. 139-142.
- [8] A. I. A. Cunha, M. C. Schneider, C. Galup-Montoro, C. D. C. Caetano, and M. B. Machado, "Extraction of MOSFET effective channel length and width based on the transconductance-to-current ratio," *Proc. Workshop on Compact Modeling, Nanotech 2005*, pp. 135-138.
- [9] Bin Yu, Clement H. J. Wann, Edward D. Nowak, Kenji Noda, and Chenming Hu, "Short-channel effect improved by lateral channel-engineering in deep-submicronmeter MOSFET's," *IEEE Trans. Electron Devices*, vol. 44, no. 4, pp. 627-634, Apr. 1997.
- [10] Matthias Bucher, Christophe Lallement and Christian C. Enz, "An efficient parameter extraction methodology for the EKV MOST model," *Proc. IEEE ICMTS*, 1996, pp. 145-150.
- [11] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "Derivation of the unified charge control model and parameter extraction procedure," *Solid-State Electron.*, vol. 43, no. 3, pp. 481-485, Mar. 1999.
- [12] Anant G. Sabnis and James T. Clemens, "Characterization of electron mobility in the inverted <100> Si surface," *IEDM Tech. Dig.*, 1979, pp. 18-21.
- [13] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An MOS transistor model for analog circuit design", *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510-1519, Oct. 1998.

Problems

11. 1. (a) Using the linear approximation of the inversion charge density with the surface potential, the relationship between the pinch-off voltage and the surface potential at which the inversion charge tends to zero, and UCCM, verify that

$$\phi_s = 2\phi_F + V_C + \phi_t \ln \left(\frac{-Q'_t}{(n-1)C'_{ox}\phi_t} \right). \quad (\text{P11.1.1})$$

(b) A classical definition of the equilibrium threshold voltage is " V_{T0} is the gate voltage at which the surface potential equals $2\phi_F$." Calculate the corresponding inversion charge density for this definition.

(c) Calculate the value of the surface potential (for $V_C=0$) which corresponds to $Q'_t = -nC'_{ox}\phi_t$.

(d) Now, assume that the threshold voltage is the gate voltage for which either $\phi_s = 2\phi_F$ or $Q'_I = -nC'_{ox}\phi_I$. Determine the difference in the threshold voltages using these two definitions.

11.2. A common method employed to determine the threshold voltage is the so-called constant current (CC) method. (a) Demonstrate that, for low V_{DS} voltages, *i.e.*, $V_{DS} / \phi_I \ll 1$, the following relationship holds

$$\frac{V_{DS}}{\phi_I} \cong (q'_{IS} - q'_{ID}) \frac{q'_{IS} + 1}{q'_{IS}} \cong \frac{I_D}{2I_S q'_{IS}}. \quad (\text{P11.2.1})$$

(b) Verify that for $V_{DS} / \phi_I = \delta$, the drain current for $V_G = V_{T0}$ (or $q'_{IS} = 1$) is given by

$$\frac{I_D}{(W/L)} = 2\delta I_{SQ}. \quad (\text{P11.2.2})$$

(c) Comment on how to use the result in (P11.2.2) to determine the threshold voltage. Also comment on the use of (P11.2.2) to determine the threshold voltage of short and/or narrow channel devices.

11.3. Assume that a long-channel MOS transistor is connected as a diode. (a) Demonstrate that, for $I_D = 3I_S$, the normalized charge at the source equals 1. Comment on the following statements: (b) If $V_S = 0$, then the value measured for the gate voltage in item (a) is the equilibrium threshold voltage. (c) If $V_S \neq 0$, then keeping $I_D = 3I_S$ allows one to determine the variation of the slope factor n with V_G . (d) An interesting circuit proposed in the literature is the specific current generator, a circuit that generates the specific current I_S of a transistor using combinations of transistors with different aspect ratios and current ratios (Reading: E. M. Camacho-Galeano, C. Galup-Montoro and M. C. Schneider, "A 2-nW 1.1-V self-biased current reference in CMOS technology," IEEE Trans. Circuits and Systems II: Express Briefs, vol. 52, no. 2, pp. 61-65, February 2005). The use of a specific current generator in item (b), instead of a constant current source, as usually done, allows one to determine the threshold voltage for any temperature.

This page intentionally left blank

Chapter 12

Advanced MOSFET Models for Circuit Simulators

The modeling of MOS transistors for integrated circuit design has been driven by the needs of digital circuit simulation for many years. The trend toward mixed analog-digital chips creates a necessity for MOSFET models appropriate for analog and RF design [1]. Strong inversion used to be the prevailing MOS operation region, but as a consequence of the technological trend toward shorter channel lengths, off-state leakage constraints, and reduced supply voltages, MOS devices now often operate in the moderate and weak inversion regions [2].

Most of the MOSFET models for computer simulation, from the SPICE Level 1 model to BSIM4, have relied on approximate solutions that are only valid in particular regions of operation connected mathematically to provide continuous solutions. Because the threshold of strong inversion V_T is the key parameter in these regional models, they are also called V_T -based models. The regional approach leads to inaccuracy between regions and consequently this class of models is not accurate enough to represent the moderate inversion region [3], widely employed in low supply voltage circuits.

For the above reasons the industry is turning away from V_T -based MOSFET models. The two main approaches as candidates to replace BSIM3-4 models are inversion-charge based and surface-potential based models. This chapter provides an overview of the approaches taken by the developers of this new generation of models. A summary of the fundamental properties of advanced MOSFET models is presented in [4], a joint paper written by the developers of the main advanced models.

12.1 Surface potential- vs. inversion charge-based models

In the early 70s, when SPICE MOSFET models began to be developed, the available computing power was very limited. Consequently, simple V_T -based models were adopted in which the surface potential is a very simple function of the input voltage: it is constant for V_{GS} above V_T and it is a linear function of gate voltage below V_T . This results in separate solutions for the different operating regions of the MOSFET, requiring smoothing functions to connect the regions. Despite their limitations, such models have been successfully used for much circuit design work. BSIM4 and MOS Model 9 are modern versions of threshold voltage-based models.

Currently, there are essentially two alternative approaches to V_T -based MOSFET models, namely surface potential-based (ϕ_s -based) [5], [6], [7] and inversion charge-based (Q'_I -based) [8], [9], [10] models. In these two alternative models, the drain current and the terminal charges are indirect functions of the terminal voltages through either the surface potential or the inversion charge density.

Conventional surface potential models based on the original charge-sheet approximation of [11] do not lead to a practical result due to difficulties in introducing velocity saturation effects for short-channels and obtaining closed-form self-consistent charges for the device terminals. Practical ϕ_s -based compact models (MM11 [6], SP [7]) use linearization of the surface potential vs. inversion charge density in a similar way as Q'_I -based models do [4]. With regard to Q'_I -based models, the inversion charge density is approximated using the unified charge control model (UCCM) [9].

Briefly, ϕ_s -based and Q'_I -based models have a common background, but enough differences exist between them to motivate model developers to support one approach rather than the other. In this chapter we will, following the joint paper by model developers [4], briefly review several recent Q'_I -based and ϕ_s -based models.

After selecting an approach to solving the fundamental differential equations, the model developer faces many more challenges due to the non-ideal real devices. These include complex doping profiles and small

dimensions that lead to a variety of physical effects. Because the complete transistor model, including the different physical effects relevant to advanced technologies, is rather complex, we will limit our presentation to the core models. The reader will find abundant literature concerning most of the models reviewed in this chapter. Here, for the sake of conciseness, we will give only the main references.

12.2 Charge-based models

A practical compact model requires efficient and accurate algorithms to calculate currents, charges and derivatives. Maher and Mead [8] showed that the drain current I_D can be expressed as a very simple function of the area densities of the inversion charge at the source, Q'_{IS} , and the drain, Q'_{ID} . This important result is expressed for a long channel transistor by (3.5.3), repeated below for convenience

$$I_D = \frac{\mu_n W}{L} \left[\frac{Q'^2_{IS} - Q'^2_{ID}}{2nC'_{ox}} - \phi_t (Q'_{IS} - Q'_{ID}) \right] \quad (12.2.1)$$

where ϕ_t is the thermal voltage and n is the slope factor.

Cunha *et al.* [10] derived expressions, which were shown in Chapter 5, for the total charges and small-signal parameters as a function of the source and drain channel charge densities. Shur's group proposed a single equation for the charge densities as a function of terminal voltages [9], called the Unified Charge Control Model (UCCM). An improved version of UCCM is given by (2.4.58) repeated below

$$V_P - V_C = \phi_t \left[\frac{Q'_{IP} - Q'_I}{nC'_{ox}\phi_t} + \ln \left(\frac{Q'_I}{Q'_{IP}} \right) \right] \quad (12.2.2)$$

where V_C is the channel voltage, V_P is the pinch-off voltage and Q'_{IP} is the value of Q'_I at pinch-off.

Recently, Gummel *et al.* [12] rediscovered basically the same charge equation and proposed a charge-based model called USIM. Q'_I -based models rely on the gradual channel assumption and a linearization of the bulk and inversion charges with respect to the surface potential at a fixed

gate bias. Because the charge-sheet approximation is not essential in the case of Q'_t -based models, they can accurately represent volume inversion and, as a consequence, they are appropriate to represent multiple gates FETs [13], [14].

12.2.1 The ACM model

The initial motivation for the ACM modeling approach came from an analog design in digital CMOS technology carried out in the late 1980's. The use of the MOS gate as a linear capacitor required the calculation of the weak nonlinearities of the MOS capacitor in accumulation and strong inversion. The classical strong inversion (SI) approximation was clearly not appropriate and improved capacitive models of the MOS gate valid for moderate inversion (MI) and accumulation were therefore developed [15].

The use of the new gate capacitor model to achieve a four terminal MOS model accurate in SI and MI, was a natural step forward. The model in [16] used lengthy surface potential-based expressions for current and charges and it was not satisfactory for analog design. The necessity for a symmetrical MOSFET model to describe the series association of transistors became clear at that time [17].

An appropriate MOSFET model was finally achieved in [10]. The symmetry of the transistor with respect to source and drain was obeyed. Rigorous definitions of pinch-off and threshold voltages, essential for consistent and precise models, were given as follows.

The channel charge density for which the diffusion current equals the drift current, designated the pinch-off charge density Q'_{IP} , is given by

$$Q'_{IP} = -nC'_{ox}\phi_t. \quad (12.2.3)$$

The channel-to-substrate voltage (V_C) for which the channel charge density equals Q'_{IP} is called the pinch-off voltage V_P .

The accurate expression for the pinch-off voltage is given by (2.4.49), repeated below

$$V_P = \phi_{sa} - 2\phi_F - \phi_t \left(1 + \ln \left(\frac{n}{n-1} \right) \right) \quad (12.2.4)$$

where ϕ_{sa} is the value of the surface potential deep in weak inversion, neglecting channel charge.

The equilibrium threshold voltage V_{TO} , measured for $V_C=0$, is the gate voltage for which the channel charge density equals Q'_{IP} .

The linear approximation of V_P can be written as

$$V_P \cong \frac{V_G - V_{TO}}{n}. \quad (12.2.5)$$

Equation (12.2.5) is useful for hand analysis, but the exponential dependence of Q'_I on V_P in weak inversion precludes (12.2.5) or expressions based on approximations of the pinch-off voltage from being used for accurate modeling.

The unified charge control model (UCCM) was derived in [20] adding a new basic approximation to the model in [10]. Considering the inversion capacitance C'_i proportional to the inversion charge it follows that

$$dQ'_I \left(\frac{1}{nC'_{ox}} - \frac{\phi_t}{Q'_I} \right) = dV_C. \quad (12.2.6)$$

Integrating (12.2.6) between an arbitrary channel potential V_C and the pinch-off voltage V_P , yields the UCCM, equation (12.2.2). As shown in Section 2.4.6, the use of the UCCM in conjunction with the pinch-off charge definition, (12.2.3), and the pinch-off voltage expression in (12.2.4) gives a model of the inversion charge essentially equivalent to the conventional surface potential charge-sheet model.

The dc, ac and non-quasi-static models were developed in [18]. The long-channel drain current expression is obtained combining expressions (12.2.1) and (12.2.2). The result was given by equations (3.7.1) and (3.7.7), rewritten below as

$$I_D = I_F - I_R = I(V_G, V_S) - I(V_G, V_D) \quad (12.2.7)$$

$$V_P - V_{S(D)} = \phi_t \left[\sqrt{1 + \frac{I_{F(R)}}{I_S}} - 2 + \ln \left(\sqrt{1 + \frac{I_{F(R)}}{I_S}} - 1 \right) \right] \quad (12.2.8)$$

where I_S is the normalization current given by

$$I_S = \mu_n C'_{ox} n \frac{\phi_t^2}{2} \frac{W}{L}. \quad (12.2.9)$$

Explicit expressions for the current, charges, transconductances and the 16 capacitive coefficients valid in weak, moderate and strong inversion were made available in [18]. All transistor parameters were given as very simple (rational) functions of the inversion charge densities at the channel boundaries.

A computer-implemented version of the model, called ACM (Advanced Compact MOSFET), has been included in a circuit simulator since 1997 [19].

Table 12.1 A complete set of capacitive coefficients for the MOSFET:

$$q'_{IS(D)} = Q'_{IS(D)} / Q'_{IP} \quad \text{and} \quad \alpha = (Q'_{ID} + Q'_{IP}) / (Q'_{IS} + Q'_{IP})$$

$C_{gs} = \frac{2}{3} C_{ox} \frac{1+2\alpha}{(1+\alpha)^2} \frac{q'_{IS}}{1+q'_{IS}}$
$C_{gd} = \frac{2}{3} C_{ox} \frac{\alpha^2+2\alpha}{(1+\alpha)^2} \frac{q'_{ID}}{1+q'_{ID}}$
$C_{bs(d)} = (n-1)C_{gs(d)}$
$C_{bg} = C_{gb} = \frac{n-1}{n} (C_{ox} - C_{gs} - C_{gd})$
$C_{sd} = -\frac{4}{15} n C_{ox} \frac{\alpha+3\alpha^2+\alpha^3}{(1+\alpha)^3} \frac{q'_{ID}}{1+q'_{ID}}$
$C_{ds} = -\frac{4}{15} n C_{ox} \frac{1+3\alpha+\alpha^2}{(1+\alpha)^3} \frac{q'_{IS}}{1+q'_{IS}}$
$C_{dg} - C_{gd} = C_m = (C_{sd} - C_{ds}) / n$

ACM has a hierarchical structure that facilitates the inclusion of different phenomena into the model. Because of its very simple expression for the derivative of the channel charge density (12.2.6), ACM was the first, and is still the only model, to furnish simple explicit expressions for all the intrinsic capacitive coefficients (given in Table 5.3 and repeated in Table 12.1).

Parameters of the ACM model can be easily extracted, as shown in [20]. Recently, unified $1/f$ noise and mismatch models were presented in [21] and [22], respectively.

12.2.2 The EKV model

The EKV model was initially developed for the design of very low-power analog ICs, in which transistors operate not only in strong inversion but also in weak and moderate inversion. The EKV model [2] was the first to introduce single-piece analytical expressions for the current, transconductances, intrinsic capacitances, non-quasi-static transadmittances, and noise valid in weak, moderate and strong inversion and from the linear to the saturation region.

The EKV model was obtained from the following approach: the weak and strong inversion asymptotes are first derived; the relevant variables (currents, conductances, etc.) are normalized and finally linked using appropriate interpolation functions. In contrast to most MOSFET models, the EKV model exploited the inherent symmetry of the MOSFET by referring all the terminal voltages to the substrate. As a consequence of the symmetry of the device, the drain current I_D is expressed as the difference between a forward component I_F and a reverse component I_R .

$$I_D = I_F - I_R. \quad (12.2.10)$$

Among the important results in [2], the following are worth highlighting.

The normalized forward and reverse currents are defined as $i_f = I_F/I_{spec}$ and $i_r = I_R/I_{spec}$, respectively. I_{spec} is the specific current defined as

$$I_{spec} = 2\mu_n C'_{ox} n\phi_t^2 \frac{W}{L}. \quad (12.2.11)$$

Observe that I_{spec} in EKV is four times larger than I_S in ACM.

The inversion charge density Q'_I is controlled by the voltage difference $V_P - V_C$, where V_P is the pinch-off voltage and V_C is the channel voltage (difference between the quasi-Fermi potentials of the carriers).

The normalized currents are given in terms of the single variable function G as

$$\frac{I_{F(R)}}{I_{spec}} = i_{f(r)} = G(V_P - V_{S(D)}). \quad (12.2.12)$$

EKV evolved into a charge-based formulation when the interpolation function for the continuous g_m/I_D characteristic was substituted by a physics-based expression covering weak to strong inversion [23]. A detailed derivation of the current inversion charge modeling in EKV can be found in [24].

The main equations of ACM and EKV are similar, but there are some differences concerning the pinch-off potential and the slope factor definitions that have some impact on the precision of the inversion charge modeling and consequently deserve some comments. The EKV model uses some empirical parameters to fine tune the inversion charge model that are not necessary in the ACM model.

The pinch-off voltage is defined as the channel voltage for which the inversion charge becomes zero under the strong inversion assumption. The pinch-off voltage in EKV is then computed as¹

$$V_P = \phi_{sa} - 2\phi_F - m\phi_t \quad (12.2.13)$$

where m can be considered as a constant, typically ranging between 1 and 4 [24].

In the EKV model, two slope factors are defined, n_q and n_v . n_q is called the inversion charge linearization factor and is given by

$$n_q = 1 + \frac{\gamma}{\sqrt{2\phi_F} + \sqrt{\phi_{sa}}}. \quad (12.2.14)$$

n_q can be interpreted as a secant between two points on the Q'_I vs. ϕ_s curve. n_v is defined in terms of the derivative of the pinch-off voltage as

¹In the EKV documentation, ϕ_{sa} is called the pinch-off surface potential (ϕ_p).

$$n_v = \left[\frac{dV_P}{dV_G} \right]^{-1} = 1 + \frac{\gamma}{2\sqrt{V_P + 2\phi_F + m\phi_t}}. \quad (12.2.15)$$

Sometimes the nuance between n_q and n_v is disregarded ($n_q \cong n_v \cong n$).

Finally, there is a small difference between ACM and EKV models concerning the value of the inversion charge at pinch-off, *i.e.*, when $V_P - V_{S(D)} = 0$.

The normalization charge is defined in the EKV model as

$$Q'_{spec} = -2n_q C'_{ox} \phi_t. \quad (12.2.16)$$

The normalized charges $q_{s(d)} = Q'_{IS(D)} / Q'_{spec}$ are dependent on the normalized terminal voltages according to [24]

$$v_p - v_{s(d)} = 2q_{s(d)} + \ln q_{s(d)}, \quad (12.2.17)$$

where the voltages are normalized with respect to the thermal voltage. The inversion charge for $v_p - v_{s(d)} = 0$ is

$$Q'_I(v_p = v_{s(d)}) = 0.852(-n_q C'_{ox} \phi_t) \quad (12.2.18)$$

while in ACM the inversion charge at pinch-off is equal to the normalization charge $-nC'_{ox}\phi_t$.

The EKV charge-based model has now evolved into a full-featured compact model that includes all the major effects that have to be accounted for in deep submicron CMOS technologies (for details, visit <http://legwww.epfl.ch/ekv>).

12.2.3 The BSIM5 model

The BSIM5 model was introduced in [25]. It uses the standard charge-based expression for the current given by Eq. (12.2.1). The inversion charge density is calculated through the UCCM using a linearized expression of the pinch-off voltage in terms of the gate voltage. In the literature available on BSIM5 there is some ambiguity concerning the exact expression used for the slope factor n . Poly-depletion and quantum effects are handled by using correction terms for the slope factor and the effective bulk potential. Velocity saturation, velocity overshoot and

source-velocity limits are modeled in a unified way. An interesting feature of BSIM5 [26] is the use of a simple empirical expression for the bulk charge valid from accumulation to strong inversion (see problem 12.1).

12.3 Surface potential models

The development of ϕ_s -based models for circuit simulation began soon after the development of the charge-sheet theory in 1978-79. In [27] short- and narrow-channel effects were included in the surface potential model of the drain current. Other surface potential models were developed [28], [29], but until recently ϕ_s -based models were considered too complicated for compact modeling.

Applying the gradual channel and charge-sheet assumptions yields the implicit relation for the surface potential presented in Section 2.4.2 and repeated below for convenience

$$(V_G - V_{FB} - \phi_s)^2 = \gamma^2 \phi_t [e^{-\phi_s/\phi_t} + \phi_s/\phi_t - 1 + e^{-(2\phi_F + V_C)/\phi_t} (e^{\phi_s/\phi_t} - \phi_s/\phi_t - 1)]. \quad (12.3.1)$$

The accurate solution of (12.3.1), once a big issue, is no longer a problem; in fact, the surface potential can be calculated within machine tolerances and with no significant burden in relation to the overall compact model calculations [30].

After determining the surface potential, the charge-sheet approximation allows a straightforward calculation of the charge densities. The charge-sheet approximation neglects the potential drop across the inversion layer for the calculation of the bulk charge density Q'_B . According to the charge-sheet approximation, Q'_B is given by

$$Q'_B = -\text{sign}(\phi_s) C'_{ox} \gamma \sqrt{\phi_s + \phi_t (e^{-\phi_s/\phi_t} - 1)}. \quad (12.3.2)$$

Expression (12.3.2) gives a continuous model from accumulation through depletion to inversion, unlike the simplified form (12.3.3), which is invalid in accumulation. The even simpler form (12.3.4), often

employed for calculations of Q'_I , is invalid even in weak accumulation as can be seen in Fig. 12.1.

$$Q'_{B1} = -C'_{ox} \gamma \sqrt{\phi_s - \phi_t} \quad (12.3.3)$$

$$Q'_{B2} = -C'_{ox} \gamma \sqrt{\phi_s}. \quad (12.3.4)$$

The inversion charge density Q'_I is given by (2.4.33), repeated below

$$Q'_I = -C'_{ox} \left(V_G - V_{FB} - \phi_s + \frac{Q'_B}{C'_{ox}} \right). \quad (12.3.5)$$

Finally, the gate electrode charge per unit area is given by

$$Q'_G = -Q'_B - Q'_I = C'_{ox} (V_G - V_{FB} - \phi_s). \quad (12.3.6)$$

Summarizing, for ϕ_s -based models, one can calculate ϕ_s iteratively using (12.3.1) and the resultant value is used to calculate Q'_B , Q'_I and, Q'_G from (12.3.2), (12.3.5) and (12.3.6), respectively.

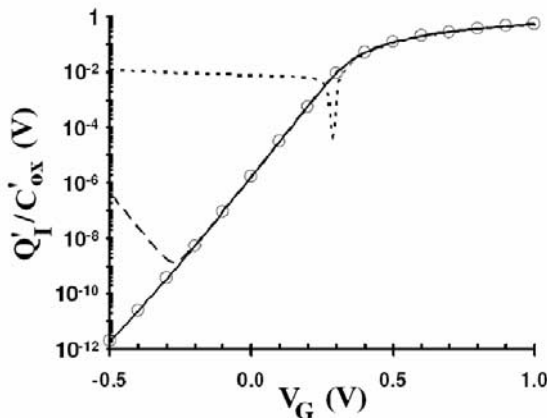


Fig. 12.1 Inversion charge density calculated with (12.3.2), full line, (12.3.3), dashed line, and (12.3.4), dotted line. Circles are results from full numerical solutions of the Poisson equation. (After [4].)

12.3.1 The HiSIM model

HiSIM (Hiroshima-University STARC IGFET Model) calculates the potentials by solving the Poisson equation iteratively at both the source and drain sides. An accuracy of 10 pV has been achieved with faster simulation times than some threshold voltage-based models [5]. Such an extreme accuracy in the surface potential calculations has turned out to be absolutely necessary for maintaining sufficiently accurate solutions for transcapacitance values as well as for achieving stable circuit simulations.

HiSIM is based on the charge-sheet approximation, as are other surface potential models, but avoids the inversion charge (Q'_i) vs. surface potential linearization (ϕ_s) employed in MM11 and SP models. The basic drain current law (for constant mobility) in HiSIM is given by equations (3.4.8), (3.4.9) and (3.4.10), repeated below for convenience

$$I_D = I_{drift} + I_{diff} \quad (12.3.7)$$

$$I_{drift} = \mu_n \frac{W}{L} C'_{ox} \left\{ (V_G - V_{FB})(\phi_{sL} - \phi_{s0}) - \frac{1}{2}(\phi_{sL}^2 - \phi_{s0}^2) - \frac{2}{3} \mathcal{N}[(\phi_{sL} - \phi_t)^{3/2} - (\phi_{s0} - \phi_t)^{3/2}] \right\} \quad (12.3.8)$$

$$I_{diff} = \mu_n \frac{W}{L} C'_{ox} \phi_t \{ (\phi_{sL} - \phi_{s0}) + \mathcal{N}[(\phi_{sL} - \phi_t)^{1/2} - (\phi_{s0} - \phi_t)^{1/2}] \}. \quad (12.3.9)$$

Without linearization of Q'_i vs. ϕ_s , the model equations, particularly those for the intrinsic charges, are complicated. As an example, the total depletion charge for $V_{BS}=0$ is given by

$$\begin{aligned}
 Q_B = & -\frac{\mu_n W^2}{I_D} \left\{ C'_{ox} q N_{sub} L_D \frac{2\sqrt{2}}{3\beta} (V'_G - \phi_s)(\beta\phi_s - 1)^{3/2} \right\}_{\phi_{s0}}^{\phi_{sL}} \\
 & + C'_{ox} q N_{sub} L_D \frac{2\sqrt{2}}{3\beta^2} \left[\frac{2}{5} (\beta\phi_s - 1)^{5/2} + (\beta\phi_s - 1)^{3/2} \right]_{\phi_{s0}}^{\phi_{sL}} \\
 & + \left(q N_{sub} L_D \sqrt{2} \right)^2 \left[\frac{\phi_s}{2} (3 - \beta\phi_s) \right]_{\phi_{s0}}^{\phi_{sL}} \quad (12.3.10)
 \end{aligned}$$

where N_{sub} is the substrate doping, L_D is the extrinsic Debye length, β is the reciprocal of the thermal voltage, the other symbols have their usual meaning, and V'_G , given below, is a shifted gate voltage that accounts for short-channel effects.

In a surface potential model, short-channel effects can no longer be accounted for by simply adding new terms to the expression for the threshold voltage. This problem has been gradually solved in the modern ϕ_s -based models by accumulating the necessary experience and trying different approaches. In both HiSIM and SP models, the bias and geometry-dependent lateral-gradient factor originally introduced in [31] [31] are employed. In this approach [5], the gate-to-source voltage V_{gs} is shifted by a value that depends on the lateral electrical field F_{yy} as shown below

$$\begin{aligned}
 V'_G &= V_{gs} + \Delta V'_G - V_{FB} \\
 \Delta V'_G &= \frac{\epsilon_s}{C'_{ox}} \sqrt{\frac{2\epsilon_s}{qN_{sub}} [\phi_s(y) - V_{bs} - \phi_t] F_{yy}} \quad (12.3.11) \\
 F_{yy} &= \frac{dF_y}{dy}.
 \end{aligned}$$

The gradient of the lateral electrical field F_{yy} is assumed to be independent of position owing to a parabolic approximation of the electrostatic potential along the channel. F_{yy} is extracted from measured threshold voltage vs. bias characteristics.

12.3.2 MOS model 11

MOS Model 11 (MM11) [6] is the successor to MOS Model 9 [32]. Its development, started in 1994, is aimed at fulfilling the requirements for a model that satisfies the accuracy demands of analog and RF circuit design but with a computational complexity which allows its application in digital designs.

To obtain efficient expressions for model outputs, several approximations were developed mainly based on a linearization of the inversion charge as a function of ϕ_s . In MM11, a linearization is performed around the average of the source and drain surface potentials [6], which results in simpler expressions without loss of accuracy.

The basic drain current expression in MM11 is the same as that given in equation (3.4.16), repeated below for convenience

$$I_D = -\frac{W}{L} \mu_n (\bar{Q}'_I - n_e C'_{ox} \phi_t) (\phi_{sL} - \phi_{s0}). \quad (12.3.12)$$

In the MM11 documentation, (12.3.12) is written as²

$$I_D = -\beta \frac{\bar{Q}^*_{inv}}{C'_{ox}} \Delta\psi \quad (12.3.13)$$

where $\beta = \mu_n C'_{ox} \frac{W}{L}$, $\bar{Q}^*_{inv} = \bar{Q}'_I - n_e C'_{ox} \phi_t$, $\Delta\psi = (\phi_{sL} - \phi_{s0})$ and C'_{ox} is the gate oxide capacitance per unit area. \bar{Q}'_I , the “average” charge density and the effective capacitance $n_e C'_{ox}$ are calculated at the average surface potential $(\phi_{sL} + \phi_{s0})/2$.

In addition, the MM11 approach also ensures that the model symmetry with respect to source-drain interchange is maintained. Note that this approach is not unlike the symmetric linearization method used in SP [7], which made it easy to merge MM11 and SP into one model called PSP.

For the description of realistic devices, however, the model has to be extended with an accurate description of mobility effects and conductance effects. In MM11, these effects have been added with a particular emphasis on distortion modeling.

²In the MM11 manual, C'_{ox} is written as C_{ox}

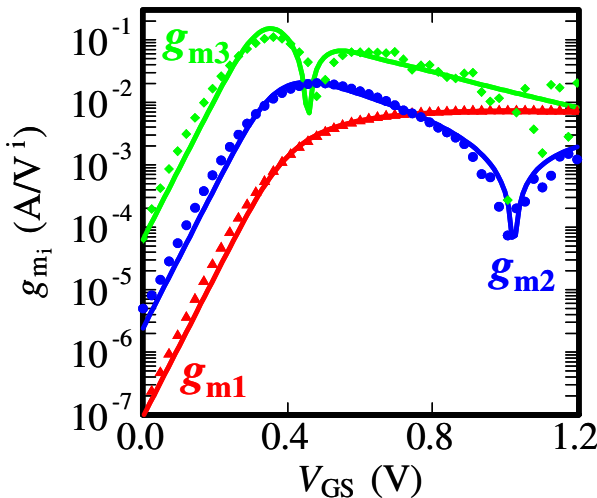


Fig. 12.2 MM11 gives an accurate description of distortion behavior. Measured (symbols) and modeled (lines) higher-order derivatives $g_{mi} (= \partial^i I_D / \partial V_{GS}^i)$ as a function of gate bias V_{GS} for $W/L=10\mu\text{m}/0.1\mu\text{m}$ device (n -MOS, $V_{DS}=0.73\text{V}$, $V_{SB}=0\text{V}$). (After [4].)

For an accurate description of distortion, the model should accurately describe the drain current and its higher-order derivatives (up to at least 3rd-order). MM11 was specially developed for this purpose. MM11 contains improved expressions for mobility reduction, velocity saturation and various conductance effects [6]. The distortion modeling of MM11 has been extensively tested on various MOSFET technologies, and it gives an accurate description of modern CMOS technologies, as seen in Fig. 12.2.

For modern CMOS technologies several physical effects that did not affect circuit design in the past have become important. All these effects should be described by the compact MOS model. MM11 includes an accurate description of all important physical effects, such as poly-depletion, quantum-mechanical effects, the effect of pocket implants, gate tunneling current, and gate-induced drain leakage [6].

The total source and drain charges for the dynamic modeling of the transistor in MM11 are given by expressions equivalent to (5.2.20) and (5.2.21), repeated below

$$Q_D = WL \left[\frac{6Q_R'^3 + 12Q_F'Q_R'^2 + 8Q_F'^2Q_R' + 4Q_F'^3}{15(Q_F' + Q_R')^2} + \frac{n}{2}C_{ox}'\phi_t \right] \quad (12.3.14)$$

$$Q_S = WL \left[\frac{6Q_F'^3 + 12Q_R'Q_F'^2 + 8Q_R'^2Q_F' + 4Q_R'^3}{15(Q_F' + Q_R')^2} + \frac{n}{2}C_{ox}'\phi_t \right]. \quad (12.3.15)$$

Rewriting (12.3.14) and (12.3.15) changing the notation as indicated below

$$-nC_{ox}' \rightarrow C_{inv} \quad (12.3.16)$$

$$Q_{F(R)}' = Q_{IS(D)}' - nC_{ox}'\phi_t \rightarrow Q_{inv0(L)} + \phi_t C_{inv} = Q_{inv0(L)}^*$$

the charge expressions as written in MM11 documentation are obtained. Clearly, the total charge expressions used in MM11 use the inversion charge density as the key variable and not the surface potential. If, concerning the input equation, MM11 is surface potential-based, concerning the output equations it is a mixed model. The current is given by (12.3.13) which combines surface potential and charge while the total charges (12.3.14) and (12.3.15), and even the $1/f$ noise expression, are given in terms of the inversion charge densities at the channel ends.

12.3.3 The SP model

The main characteristics of the SP model can be summarized as follows [7]: ϕ_s -based, substrate-referenced with analytic (non-iterative) computation of ϕ_s from accumulation to inversion, symmetric with respect to source-drain interchange, and physics-based modeling of small-geometry effects via lateral field gradient, as in HiSIM.

The search for an approximate analytical solution of the surface potential master equation was initiated in [28] using smoothing functions to connect the asymptotic solutions in subthreshold and strong inversion. However, even the best approximations of this type [33] have an accuracy of 2–3 mV which is not sufficient for the modeling of transconductances and transcapacitances. By using a different approach, SP obtains an approximate solution of (12.3.1) with accuracy better than

1 nV [34]. As shown in Fig. 12.3, the accuracy of the SP algorithm is the same as that in [34], but now the accumulation region is included.

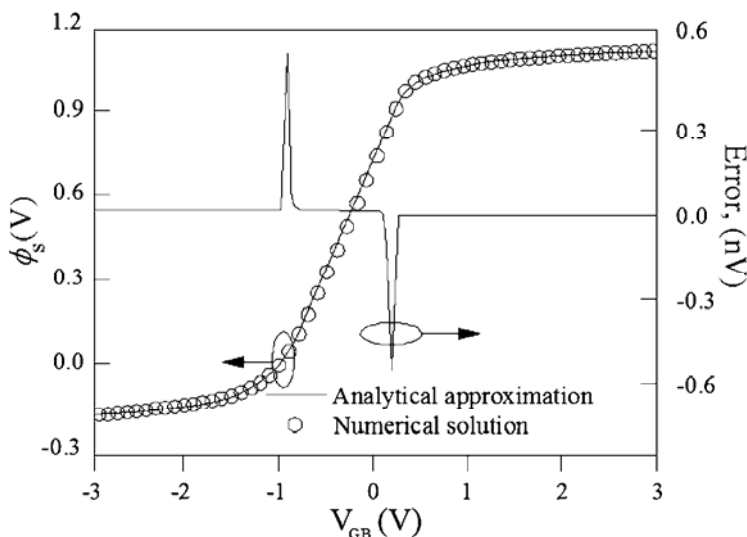


Fig.12.3 Comparison of the analytical approximation of the surface potential in the channel area with numerical solution: $N_{\text{sub}}=5 \times 10^{17} \text{ cm}^{-3}$, $t_{\text{ox}}=25 \text{ \AA}$, $V_{\text{FB}}=-1 \text{ V}$, $T=300 \text{ K}$. (After [7].)

Bibliography

- [1] Yannis Tsividis and Guido Masetti, "Problems in precision modeling of the MOS transistor for analog applications," *IEEE Trans. Computer-Aided Design*, vol. 3, no. 1, pp. 72-79, Jan. 1984.
- [2] Christian C. Enz, François Krummenacher and Eric A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *J. Analog Integr. Circuits Signal Process.*, vol. 8, pp. 83-114, July 1995.
- [3] Y. Tsividis, "Moderate inversion in MOS devices," *Solid-State Electron*, vol. 25, no. 11, pp. 1099-1104, Nov. 1982. See also Erratum, *ibid.*, vol. 26, no. 8, p. 823, Aug. 1983.
- [4] Josef Watts, Colin McAndrew, Christian Enz, Carlos Galup-Montoro, Gennady Gildenblat, Chenming Hu, Ronald van Langevelde, Mitiko Miura-Mattausch, Rafael Rios, Chih-Tang Sah, "Advanced compact models for MOSFETs," *Proc. Workshop on Compact Modeling, Nanotech 2005*, pp. 3-12.
- [5] Mitiko Miura-Mattausch, Uwe Feldmann, Alexander Rahm, Michael Bollu, and Dominique Savignac, "Unified complete MOSFET model for analysis of digital

- and analog circuits," *IEEE Trans. Computer-Aided Design*, vol. 15, no. 1, pp. 1-7, Jan. 1996.
- [6] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "Physical background of MOS Model 11," Nat. Lab. Unclassified Report 2003/00239. [On line] Available: http://www.semiconductors.philips.com/Philips_Models/
- [7] Gennady Gildenblat, Hailing Wang, Ten-Lon Chen, Xin Gu, and Xiaowen Cai, "SP: An advanced surface-potential-based compact MOSFET model," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1394-1406, Sep. 2004.
- [8] Mary Ann Maher and Carver A. Mead, "A physical charge-controlled model for MOS transistors," in *Advanced Research in VLSI*, P. Losleben (ed.), MIT Press, Cambridge, MA, pp. 211-229, 1987.
- [9] Young Hee Byun, Kwyro Lee, and Michael Shur, "Unified charge control model and subthreshold current in heterostructure field effect transistors," *IEEE Electron Device Lett.*, vol. 11, no. 1, pp. 50-53, Jan. 1990.
- [10] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electron.*, vol. 38, no. 11, pp. 1945-1952, Nov. 1995.
- [11] J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electron.*, vol. 21, no. 2, pp. 345-355, Feb. 1978.
- [12] Hermann K. Gummel and Kumud Singhal, "Inversion charge modeling," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1585-1593, Aug. 2001.
- [13] Benjamin Iñíguez, David Jiménez, Jaume Roig, Hamdy A. Hamid, Lluís F. Marsal, and Josep Pallarès, "Explicit continuous model for long-channel undoped surrounding gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 52, no. 8, pp. 1868-1873, Aug. 2005.
- [14] Giorgio Mugnaini and Giuseppe Iannaccone, "Physics-based compact model of nanoscale MOSFETs-Part I: transition from drift-diffusion to ballistic transport", *IEEE Trans. Electron Devices*, vol. 52, no. 8, pp. 1795-1801, Aug. 2005.
- [15] Alexandre Ternes Behr, Márcio Cherem Schneider, Sidnei Noceti Filho, and Carlos Galup Montoro, "Harmonic distortion caused by capacitors implemented with MOSFET gates," *IEEE J. Solid-State Circuits*, vol. 27, no. 10, pp. 1470-1475, Oct. 1992.
- [16] A. I. A. Cunha, R. T. Gonçalves, M. C. Schneider, and C. G. Montoro, "An accurate and explicit physical model of the MOS transistor", *Proc. of the 8th Brazilian Conference on Microelectronics*, 1993, Campinas, Brazil, pp. V7-V12.
- [17] Carlos Galup-Montoro, Márcio C. Schneider, and Itamar J. B. Loss, "Series-parallel association of FET's for high gain and high frequency applications," *IEEE J. Solid-State Circuits*, vol. 29, no. 9, pp. 1094-1101, Sep. 1994.
- [18] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510-1519, Oct. 1998.
- [19] Application Notes in Home-page Dolphin. [Online]. Available: http://www.dolphin.fr/medal/smash/notes/acm_report.pdf

- [20] Ana Isabela Araújo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro, "Derivation of the unified charge control model and parameter extraction procedure," *Solid-State Electron.*, vol. 43, no. 3, pp. 481-485, Mar. 1999.
- [21] Alfredo Arnaud and Carlos Galup-Montoro, "A compact model for flicker noise in MOS transistors for analog circuit design," *IEEE Trans. Electron Devices*, vol. 50, no. 8, pp. 1815-1818, Aug. 2003.
- [22] Carlos Galup-Montoro, Márcio C. Schneider, Hamilton Klimach, and Alfredo Arnaud, "A compact model of MOSFET mismatch for circuit design," *IEEE J. Solid-State Circuits*, vol. 40, no. 8, pp. 1649-1657, Aug. 2005.
- [23] Matthias Bucher, *Analytical MOS Transistor Modelling for Analog Circuit Simulation*, PhD Thesis, EPFL, Lausanne, Switzerland, 1999.
- [24] Jean-Michel Sallese, Matthias Bucher, François Krummenacher, and Pierre Fazan, "Inversion charge linearization in MOSFET modeling and rigorous derivation of the EKV compact model," *Solid-State Electron.* vol. 47, no. 4, pp. 677-683, Apr. 2003.
- [25] Jin He, Xuemei Xi, Mansun Chan, Ali Niknejad, and Chenming Hu, "An advanced surface potential-plus MOSFET model," *Proc. Workshop on Compact Modeling, Nanotech 2003*, pp. 262-265.
- [26] Ali Niknejad, BSIM5 presentation. [Online]. Available: http://www.eigroup.org/cmc/next_gen_cmos/bsim5latest.pdf
- [27] P. P. Guebels and F. Van de Wiele, "A charge sheet model for small geometry MOSFET's," *IEDM Tech. Dig.*, 1981, pp. 211-214.
- [28] Claudio Turchetti and Guido Masetti, "A CAD-oriented analytical MOSFET model for high-accuracy applications," *IEEE Trans. Computer-Aided Design*, vol. 3, no. 2, pp. 117-122, Apr. 1984.
- [29] A. R. Boothroyd, Stan W. Tarasewicz, and Cezary Slaby, "MISNAN - A physically based continuous MOSFET model for CAD applications," *IEEE Trans. Computer-Aided Design*, vol. 10, no. 12, pp. 1512-1529, Dec. 1991.
- [30] Rafael Rios, Sivakumar Mudanai, Wei-Kai Shih, and Paul Packan, "An efficient surface potential solution algorithm for compact MOSFET models," *IEDM Tech. Dig.*, 2004, pp. 755-758.
- [31] Mitiko Miura-Mattausch and Hermann Jacobs, "Analytical model for circuit simulation with quarter micron metal oxide semiconductor field effect transistors: subthreshold characteristics," *Jap. J. App. Phys.*, vol. 29, no. 12, pp. 2279-2282, Dec. 1990.
- [32] R. M. D. A. Velghe, D. B. M. Klaassen, F. M. Klaassen, "MOS Model 9," NL-UR 003/94, Philips Electronics N.V. [Online]. Available: http://www.semiconductors.philips.com/Philips_Models/
- [33] Ronald van Langevelde and François M. Klaassen, "An explicit surface-potential-based MOSFET model for circuit simulation," *Solid-State Electron.*, vol. 44, no. 3, pp. 409-418, Mar. 2000.
- [34] T.-L. Chen and G. Gildenblat, "Analytical approximation for the MOSFET surface potential," *Solid-State Electron.*, vol. 45, no. 2, pp. 335-339, Feb. 2001.

Problems

12.1. (a) Verify that, in inversion, the bulk charge density can be written as

$$Q'_B \cong -\gamma C'_{ox} \sqrt{\phi_s} \cong -\gamma C'_{ox} \left(\sqrt{V_G - V_{FB} + \frac{Q'_I}{C'_{ox}} + \frac{\gamma^2}{4} - \frac{\gamma}{2}} \right).$$

(b) What is the asymptotic value of the bulk charge deep in accumulation?

(c) Combine the results of items (a) and (b) to demonstrate that the smoothing function

$$Q'_B \cong -C'_{ox} \frac{V_G - V_{FB} + \frac{Q'_I}{C'_{ox}}}{\frac{1}{2} + \sqrt{\frac{V_G - V_{FB} + \frac{Q'_I}{C'_{ox}}}{2\gamma^2} + \frac{1}{4} + \frac{1}{2} \left[\sqrt{\left(\frac{V_G - V_{FB} + \frac{Q'_I}{C'_{ox}}}{\gamma^2} \right)^2 + \delta} \right]}}$$

with $\delta \ll 1$ represents correctly the asymptotic behavior of the bulk charge. This result for the bulk charge was previously presented on page 19 of reference [26].

Appendix A

Electrostatics in One Dimension

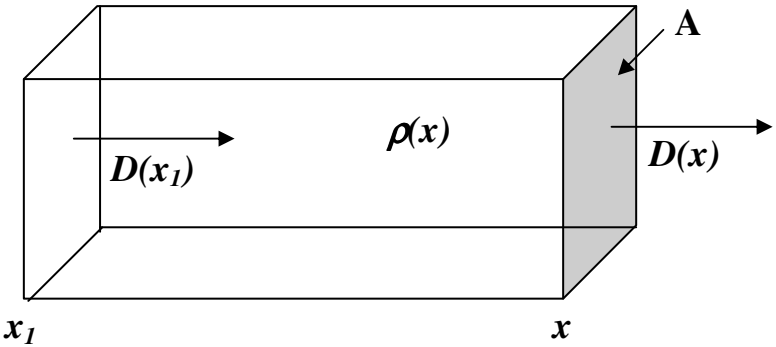


Fig. A.1 Rectangular Gaussian box parallel to a one-dimensional electric field. D is the electric displacement and ρ is the charge density.

The outward flow, or flux, of a field through an element of surface is equal to the outward component of the field perpendicular to the surface times the area of the surface. Gauss' law states that the flux of the electric displacement D flowing out of a closed surface is equal to the net charge Q enclosed within the surface:

$$\text{The flux of } D \text{ through any closed surface} = Q. \quad (\text{A.1})$$

For the cases of interest in this book, the relation between the electric field F and the electric displacement D is given by

$$D = \epsilon F \quad (\text{A.2})$$

where ϵ is the permittivity of the material, assumed to be a scalar.

Let us consider in the case of a one-dimensional vector field a rectangular Gaussian box parallel to the field direction as shown in

Fig. A1. Since the field runs parallel to the lateral sides of the box, there is no flux contribution from the lateral sides and Gauss' law reduces to

$$AD(x) - AD(x_1) = A \int_{x_1}^x \rho(x) dx \quad (\text{A.3})$$

where A is the area of the lateral side of the box as indicated in Fig. A.1 and ρ is the volumetric charge density. Considering a constant permittivity ϵ across the sample and canceling the common term A gives

$$F(x) - F(x_1) = \frac{1}{\epsilon} \int_{x_1}^x \rho(x) dx. \quad (\text{A.4})$$

Let us consider the case (common in bulk MOS modeling) where the electric field is zero in x_1 .

Equation (A.4) reduces to

$$F(x) = \frac{1}{\epsilon} \int_{x_1}^x \rho(x) dx = \frac{Q'}{\epsilon} \quad (\text{A.5})$$

where $Q' = Q/A$ is the charge per unit area or areal charge density.

The differential form of Gauss's law is obtained taking the derivative of (A.4) with respect to x

$$\frac{dF}{dx} = \frac{\rho}{\epsilon}. \quad (\text{A.6})$$

The electrostatic field is given by the gradient of the electric potential ϕ as

$$\frac{d\phi}{dx} = -F. \quad (\text{A.7})$$

Combining (A.6) and (A.7) yields the Poisson equation

$$\frac{d^2\phi}{dx^2} = -\frac{\rho}{\epsilon}. \quad (\text{A.8})$$

Electrostatics in one dimension reduces to the study of the solutions of (A.8). In other terms, the Maxwell equations for the electrostatic field are equivalent to the Poisson equation (A.8).

To finish, let us consider two materials with different permittivities ϵ_1 and ϵ_2 as indicated below

$$\begin{aligned}\epsilon &= \epsilon_1 \text{ for } x \leq x_0 \\ \epsilon &= \epsilon_2 \text{ for } x > x_0.\end{aligned}\tag{A.9}$$

Considering that there is no charge sheet of zero thickness at x_0 , Gauss' law (A.3) gives the boundary condition

$$\lim_{x \rightarrow x_0^+} \epsilon_2 F(x) = \epsilon_1 F(x_0).\tag{A.10}$$

Bibliography

- [1] Richard P. Feynman, Robert B. Leighton, Matthew L. Sands, *The Feynman Lectures on Physics*, Addison-Wesley, Reading, Mass., 1964.
- [2] Yannis Tsividis, *Operation and modeling of the MOS transistor*, 2nd ed., McGraw-Hill, Boston, 1999.

This page intentionally left blank

Appendix B

Electrostatics in Semiconductors

To calculate the charge density ρ inside a semiconductor let us recall some basics of semiconductors. Two types of foreign atoms called impurities are introduced into semiconductors in order to modify the electron and hole carrier densities. This process, called doping, is the most important feature of semiconductors; the fabrication of semiconductor devices is only possible because of it. Donor atoms can donate an electron to the crystal lattice and acceptor atoms can accept an electron from the valence band producing a hole. Donor materials come from group V of the periodic table; the most commonly used materials for n-type silicon are phosphorus and arsenic. Acceptors are elements from group III of the periodic table; the most common acceptor material is boron. Because the dopant atoms are originally neutral, donors become positively ionized when donating an electron while acceptors become negatively ionized when accepting an electron. As a consequence, in the general case, four types of charge are present inside a semiconductor: the fixed positive charge of ionized donors, the fixed negative charge of ionized acceptors, the positive mobile charge of holes, and the negative mobile charge of electrons. We will consider the situations in which all donors and acceptors are ionized, *i.e.*

$$N_D = N_D^+ \quad \text{and} \quad N_A = N_A^-, \quad (\text{B.1})$$

where $(N_D^+)N_D$ and $(N_A^-)N_A$ are the (ionized) donor and acceptor atoms per unit volume. On this basis, the net positive charge density is given by

$$\rho = q(N_D - N_A + p - n). \quad (\text{B2})$$

Let us analyze the small deviations from neutrality in the simple case of a homogeneous semiconductor (p-type semiconductor with $N_D=0$) in thermal equilibrium. The carrier concentrations are given by Boltzmann's law, equations (2.1.5) and (2.1.6). Thus, the charge density in (B.2) is given by

$$\rho = q(p_0 e^{-u} - N_A - n_0 e^u), \quad (\text{B.3})$$

where u is the normalized (to thermal voltage) electrostatic potential. The origin of the potential ($u=0$) in (B.3) has been chosen to be in the neutral region, where $p_0=N_A+n_0$. For a reasonably extrinsic p-type material ($p_0 \gg n_0$) and for a small departure from neutrality ($|u| \ll 1$),

$$p_0 e^{-u} \gg n_0 e^u. \quad (\text{B.4})$$

Thus, (B.3) can be approximated by

$$\rho \cong q N_A (e^{-u} - 1) \cong -q N_A u. \quad (\text{B.5})$$

Finally, substituting (B.5) into the Poisson equation (A.8) yields

$$\frac{d^2 u}{dx^2} = \frac{q^2 N_A}{k T \epsilon_s} u. \quad (\text{B.6})$$

The solution of (B.6), which vanishes properly as $x \rightarrow +\infty$, is of the form

$$u = \text{Constant} \cdot \exp(-x/L_D) \quad (\text{B.7})$$

where

$$L_D = \sqrt{\frac{k T \epsilon_s}{q^2 N_A}} \quad (\text{B.8})$$

is the extrinsic Debye length, a measure of a characteristic distance for departure from charge neutrality.

Bibliography

- [1] Richard B. Adler, Arthur C. Smith, and Richard L. Longini, *Introduction to Semiconductor Physics*, SEEC Series, vol. 1, John Wiley & Sons, New York, 1964.

Appendix C

Drift-Diffusion Current Model

The motion of a carrier in a semiconductor is affected by the interaction of the carrier with the periodic crystal forces. This interaction is accounted for by the use of an effective mass, which differs from the free electron mass. The motion of electrons and holes in semiconductors can thus be approximately described by the effective mass approximation, which assumes that the electrons and holes behave as free particles where the influences of crystal forces are incorporated into this effective mass.

Electrons (and holes) have the thermal energy of classical free particles, that is, $kT/2$ for each degree of freedom, where k is the Boltzmann constant and T is the temperature in Kelvin.

For electrons (moving in three dimensions), the kinetic energy is

$$\frac{1}{2}m_n v_{th}^2 = \frac{3}{2}kT \quad (C.1)$$

where v_{th} is the thermal velocity and m_n is the effective mass for free (conduction-band) electrons. For electrons in silicon at $T=300$ K, $v_{th} \cong 10^7$ cm/s.

A very simple model of carrier transport or current flow in semiconductors is obtained considering the carrier mean time between collisions (or mean scattering time) τ and the mean free-path $l = v_{th} \tau$. To simplify further, we consider transport in one dimension.

When a weak electric field F is applied to a semiconductor the carriers acquire a drift velocity v_d added to their random (thermal) motion which is proportional to the applied field, the proportionality constant μ called the mobility. For electrons it holds that

$$v_d = -\mu_n F . \quad (C.2)$$

Considering that in steady state the momentum gained by the particles between collisions is lost to the lattice after each collision, the expression below holds

$$\mu_n = \frac{q\tau}{m_n} . \quad (C.3)$$

This result is as simple as possible; the mobility is proportional to the mean scattering time τ and is inversely proportional to the effective mass of the carriers.

The current density can be found through the sum of the product of the charge of each electron times its velocity over all electrons per unit volume n , which together with (C.2) gives

$$J_n = -nqv_d = nq\mu_n F . \quad (C.4)$$

For a uniform concentration of carriers the thermal agitation results in thermal noise, but in zero steady state current for $F=0$. For a non uniform concentration of carriers, a net current results from the thermal agitation of carriers, due to the net flux of carriers moving from the higher concentration regions into the lower concentration regions. This diffusion current is (with a very good approximation) proportional to the concentration gradient as follows

$$J_n = qlv_{th} \frac{dn}{dy} = qD_n \frac{dn}{dy} . \quad (C.5)$$

The proportionality constant D_n , called the diffusion coefficient, is proportional to the thermal velocity and to the mean free-path. Again this is the simplest possible model.

Drift and diffusion are different transport mechanisms, but both are related to the random thermal velocity of the carriers. To make this link explicit, let us rewrite the diffusion coefficient in terms of the mean scattering time using $l = v_{th} \tau$

$$D_n = lv_{th} = \tau v_{th}^2 . \quad (C.6)$$

Applying the theorem for the equipartition of energy to the motion in one dimension

$$\frac{1}{2}m_n v_{th}^2 = \frac{1}{2}kT \quad (C.7)$$

yields the Einstein relationship

$$D_n = \tau v_{th}^2 = \frac{\tau kT}{m_n} = \frac{kT}{q} \frac{q\tau}{m_n} = \phi_t \mu_n. \quad (C.8)$$

In the presence of a field $F = -d\phi/dy$ and of a gradient of concentration, we have a drift current and a diffusion current. In this case, the total current density at any point is the sum of the drift and diffusion components

$$J_n = qn\mu_n \left(-\frac{d\phi}{dy} \right) + qD_n \frac{dn}{dy}. \quad (C.9)$$

A similar expression holds for holes

$$J_p = qp\mu_p \left(-\frac{d\phi}{dy} \right) - qD_p \frac{dp}{dy}, \quad (C.10)$$

where the negative sign in front of the diffusion component arises from the positive charge of the hole.

The total current density is

$$J = J_n + J_p. \quad (C.11)$$

Bibliography

- [1] Richard S. Muller and Theodore I. Kamins, *Device Electronics for Integrated Circuits*, John Wiley&Sons, New York, 1977.

This page intentionally left blank

Appendix D

Continuity Equations

To derive the continuity equation for holes, let us consider a unidimensional current flow and the box shown in Fig. D.1, where J_p is the hole current density, and G_p and R_p refer to the generation and recombination rates for holes, per unit time per unit volume. The continuity equations result from the charge conservation principle.

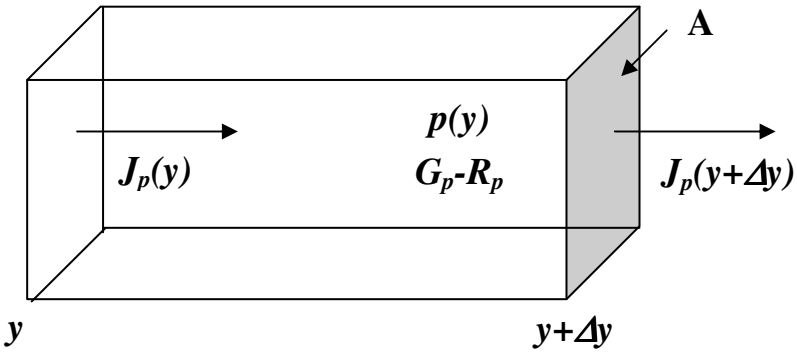


Fig. D.1 Illustration of the continuity equation for a one-dimensional flow of holes.

Conservation of holes implies that the net number of holes entering the volume of cross section A and length Δy per unit time, namely,

$$\frac{1}{q} A [J_p(y) - J_p(y + \Delta y)] \quad (\text{D.1})$$

plus the number of holes generated within the volume per unit time

$$[G_p - R_p] \Delta y A \quad (\text{D.2})$$

equals the increase in the number of holes in the volume per unit time

$$\frac{\partial}{\partial t}(pA\Delta y). \quad (\text{D.3})$$

Thus, equating the above terms, and dividing by the volume $\Delta y.A$, gives

$$-\frac{1}{q} \frac{[J_p(y+\Delta y) - J_p(y)]}{\Delta y} + G_p - R_p = \frac{\partial p}{\partial t}. \quad (\text{D.4})$$

In the limit case as $\Delta y \rightarrow 0$, we have

$$-\frac{1}{q} \frac{\partial J_p}{\partial y} + G_p - R_p = \frac{\partial p}{\partial t}. \quad (\text{D.5})$$

A similar relation holds for electrons

$$\frac{1}{q} \frac{\partial J_n}{\partial y} + G_n - R_n = \frac{\partial n}{\partial t}. \quad (\text{D.6})$$

Bibliography

- [1] Richard B. Adler, Arthur C. Smith, and Richard L. Longini, *Introduction to Semiconductor Physics*, SEEC Series, vol. 1, John Wiley & Sons, New York, 1964.

Appendix E

Basics of pn Junctions

This appendix summarizes the basic properties of the pn junctions that are relevant to the intrinsic MOS transistor. Thus, the current-voltage and capacitance-voltage characteristics of the pn junction are not addressed here. We focus on the pn junction as a contact with the inversion layer; therefore, only the electrostatics of the junction necessary to establish the boundary conditions for the inversion channel is reviewed.

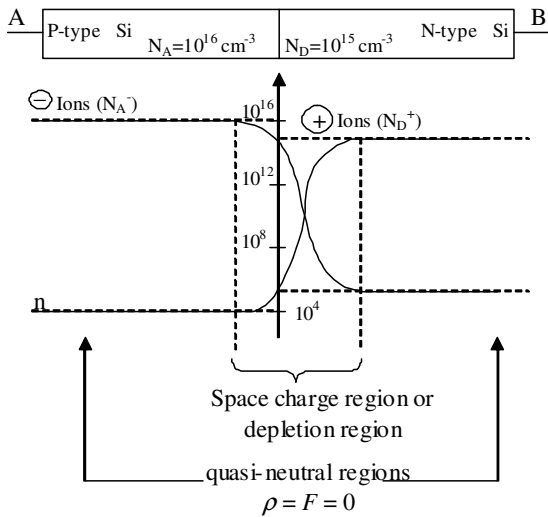


Fig. E.1 Diagram of carrier densities across a pn junction.

E.1 Electrostatics in equilibrium

A pn junction is present when a transition between p-type and n-type doping occurs in adjacent regions of a semiconductor (crystal). Junctions are called abrupt when the transition from the p-type to the n-type doping is very narrow. The doping profile and the carrier densities of an abrupt junction, with constant doping densities on either side of the junction, are shown in Fig. E.1. As a consequence of the carrier concentration differences between the p-type and n-type regions, electrons (holes) diffuse from the n (p) side to the p (n) side, leaving behind ionized donor (acceptor) atoms and, consequently, a charge imbalance and a net charge density. The space charge region near the metallurgical junction where the mobile carrier concentrations have been reduced below their values far from the junction is called the depletion region. Far from the metallurgical junction, neutrality prevails as shown in Fig. E.1. The electric field produced by the charge imbalance counteracts the diffusion current so that in thermal equilibrium the electron and hole currents are zero. Setting the total electron current to zero yields

$$J_n = qn\mu_n \left(-\frac{d\phi}{dy} \right) + qD_n \frac{dn}{dy} = 0. \quad (\text{E.1})$$

Solving for the electric field and using the Einstein relationship $D_n = \phi_t \mu_n$, yields

$$\frac{d\phi}{dy} = \phi_t \frac{1}{n} \frac{dn}{dy}. \quad (\text{E.2})$$

Integrating (E.2) between any two points gives

$$\phi_b - \phi_a = \phi_t \ln(n_b / n_a). \quad (\text{E.3})$$

Far from the junction on the p-side we have

$$n_a = n_p = n(-\infty) \cong n_i^2 / N_A, \quad (\text{E.4})$$

and far from the junction on the n-side

$$n_b = n_n = n(+\infty) \cong N_D. \quad (\text{E.5})$$

Thus, from (E.3)–(E.5) we obtain the internal potential barrier height or built-in potential of the junction

$$\phi_b - \phi_a = \phi_{bi} = \phi_t \ln \left(\frac{N_A N_D}{n_i^2} \right), \quad (\text{E.6})$$

calculated between two points far from the junction. The ideally ohmic contacts of the metal with the p-type and n-type semiconductor have contact potentials V_P and V_N , respectively, such that the potential drops around the loop (metal, p-type and n-type semiconductor) add up to zero.

E.2 Electrostatics of forward- and reverse-biased junction

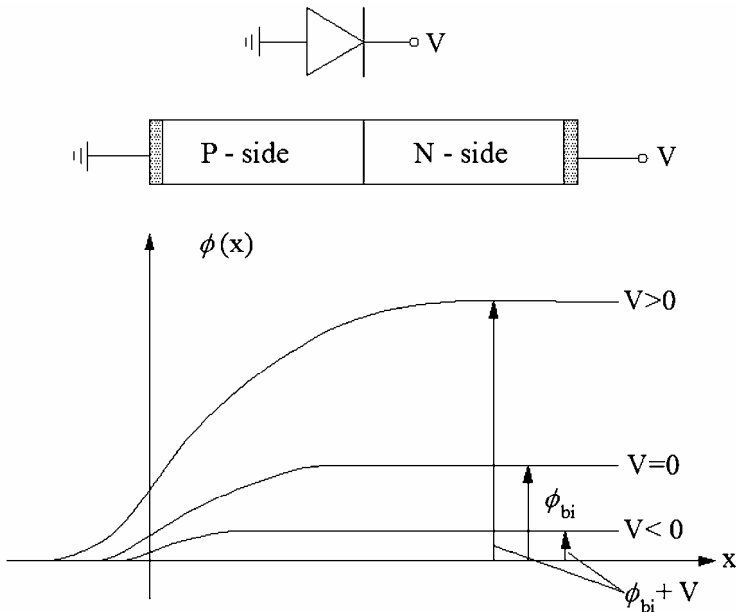


Fig. E.2 Potential barrier inside a biased junction.

In the fundamental model of the junction it is considered that the contact potentials of the ohmic contacts are unchanged under bias and also that the ohmic potential drops in the quasi-neutral n-type and p-type regions can be neglected. That is, all ohmic drops in the bulk and contact are negligible compared to the resistance presented by the semiconductor junction itself. As a consequence, all the applied bias will drop in the

depletion region. The applied voltage is (algebraically) added to the built-in potential, increasing (decreasing) the internal potential barrier for reverse (forward) bias. Thus, the barrier height for an applied bias V (considered positive for a reverse biased junction) is

$$\text{barrier height} = \phi_{bi} + V \quad (\text{E.7})$$

as shown in Fig. E.2.

E.3 Quasi-Fermi potentials in the pn junction

Let us consider a biased pn junction, as shown in Fig. E.3. The definitions of quasi-Fermi levels of expressions (2.4.1) and (2.4.2) are repeated below for convenience.

$$p = n_i e^{(\phi_p - \phi)/\phi_i} \quad (\text{E.8})$$

$$n = n_i e^{(\phi - \phi_n)/\phi_i} \quad (\text{E.9})$$

From the definition of quasi-Fermi levels given in (E.8) and (E.9), the differences $(\phi_p - \phi)$ and $(\phi - \phi_n)$ in the neutral regions far from the metallurgical junction must be the same as in equilibrium, because the majority carrier densities are (almost) the same as in equilibrium.

Given that the applied reverse (forward) voltage appears as an increase (decrease) in the potential barrier inside the semiconductor, the quasi-Fermi levels must split by the amount of the applied reverse (forward) bias, to keep the majority carrier concentrations in the neutral regions on both sides of the junction constant as shown in Fig. E.3, *i.e.*,

$$\phi_n - \phi_p = V \quad (\text{E.10})$$

where V is the voltage applied to the junction, which is positive for a reverse biased junction.

Since the hole current is carried by a high hole concentration in the p-region and by a low hole concentration in the n-region, the gradient of ϕ_p must be much greater in the n- than in the p-region, as shown in Fig. E.3. For the same reason, the variation in ϕ_n occurs mainly in the p-region.

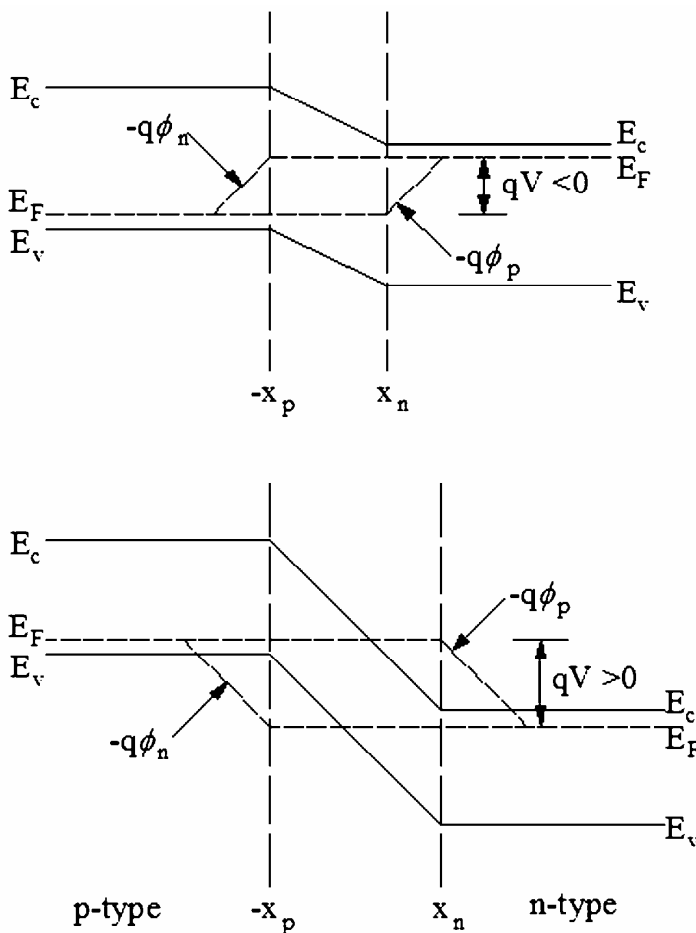


Fig. E.3 Variation in the electron and hole quasi-Fermi levels for a (a) forward biased and (b) reverse biased junction. (After [3].)

Bibliography

- [1] Gerold W. Neudeck, *The PN Junction Diode*, 2nd ed., Addison-Wesley, Reading, Mass., 1989.
- [2] William Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Co., Princeton, NJ, 1950.
- [3] Yuan Taur and Tak H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.

This page intentionally left blank

Appendix F

Hall-Shockley-Read (HSR) Statistics

Recombination of carriers plays an important role in transistor physics. Imperfections within the semiconductor crystal can introduce energy levels in the forbidden gap, usually called traps or recombination centers.

The four possible transitions of an electron between an energy level in the forbidden band and the conduction or the valence bands are indicated in Fig. F.1.

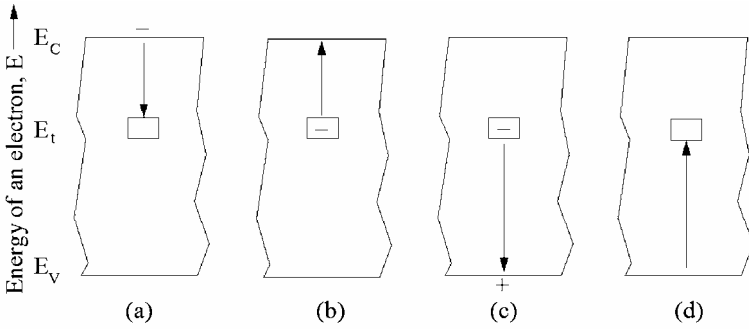


Fig. F.1 The basic processes involved in recombination by trapping: (a) electron capture, (b) electron emission, (c) hole capture, and (d) hole emission (After [2].)

The rate of electron capture r_a is proportional to the concentration of free electrons n and to the concentration of traps not occupied by electrons. Thus,

$$r_a = v_{th} \sigma_n n N_t (1 - f_{HSR}) \quad (F.1)$$

where r_a is the electron capture rate, v_{th} is the thermal velocity, σ_n is the capture cross-section for electrons, n is the electron concentration, N_t is

the trap volumetric density and f_{HSR} is the probability of occupation of a center by an electron.

The rate of electron emission r_b is proportional to the concentration of traps which are occupied by electrons. Thus,

$$r_b = e_n N_t f_{HSR} \quad (F.2)$$

where r_b is the electron emission rate and e_n is the emission probability of an electron.

The rate of capture of holes, by analogy to process (a), is given by

$$r_c = v_{th} \sigma_p p N_t f_{HSR} \quad (F.3)$$

where r_c is the hole capture rate, σ_p is the capture cross-section for holes, and p is the hole concentration.

Finally, the rate of hole emission, process (d), is

$$r_d = e_p N_t (1 - f_{HSR}) \quad (F.4)$$

where r_d is the hole emission rate and e_p is the emission probability of a hole.

In equilibrium, the occupation of the traps is given by Fermi statistics, thus HSR statistics reduces to Fermi-Dirac statistics f_{FD} as indicated below

$$f_{HSR} = f_{FD} = \frac{1}{1 + \exp[(E_t - E_F) / kT]} \quad (F.5)$$

where E_t is the energy level of the trap and E_F is the Fermi level. Considering that in equilibrium

$$r_a - r_b = r_c - r_d = 0 \quad (F.6)$$

and using Boltzmann statistics to calculate the free carrier densities

$$n = n_i \exp[(E_F - E_i) / kT] \quad p = n_i \exp[(E_i - E_F) / kT] \quad (F.7)$$

where E_i is the intrinsic Fermi level, it follows that

$$e_n = v_{th} \sigma_n n_i \exp\left(\frac{E_t - E_i}{kT}\right) \quad (F.8)$$

and

$$e_p = v_{th} \sigma_p n_i \exp\left(\frac{E_i - E_t}{kT}\right). \quad (\text{F.9})$$

Out of equilibrium, but in steady state

$$r_a - r_b = r_c - r_d (\neq 0). \quad (\text{F.10})$$

Assuming that out of equilibrium (F.8) and (F.9) continue to be valid, it follows from (F.1-4) and (F.8-9) that

$$f_{HSR} = \frac{\sigma_n n + \sigma_p n_i \exp\left(\frac{E_i - E_t}{kT}\right)}{\sigma_n [n + n_i \exp\left(\frac{E_t - E_i}{kT}\right)] + \sigma_p [p + n_i \exp\left(\frac{E_i - E_t}{kT}\right)]}. \quad (\text{F.11})$$

Substituting (F.11) in (F.1) and (F.2) and assuming that the capture cross-sections for holes and electron are equal ($\sigma_p = \sigma_n = \sigma$), one finds that the net recombination rates of electrons and holes in steady state are

$$r_a - r_b = r_c - r_d = \sigma v_{th} N_t \frac{pn - n_i^2}{p + n + 2n_i \cosh\left[\left(E_i - E_t\right)/kT\right]}. \quad (\text{F.12})$$

Bibliography

- [1] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, New York, 1967.
- [2] W. Shockley and W. T. Read Jr., "Statistics of the recombination of holes and electrons", *Phys. Rev.*, vol. 87, no.5, pp. 835-842, 1 Sep. 1952
- [3] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, no.2, p. 387, July 1952.

This page intentionally left blank

Appendix G

Interface Trap Capacitance

In equilibrium, the probability of occupancy of interface traps is given by the Fermi-Dirac distribution function which gives the probability of a given energy state being occupied by an electron,

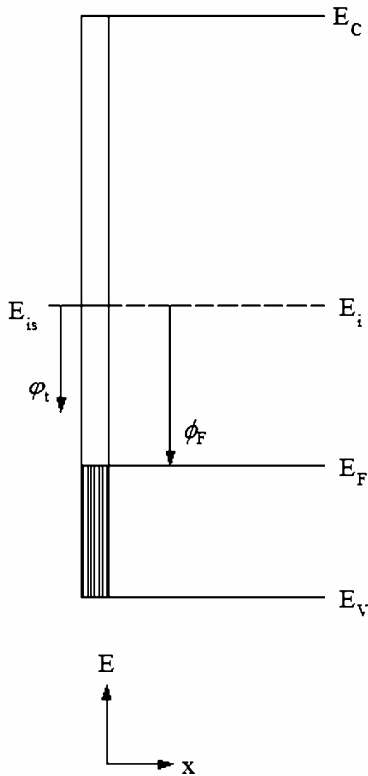


Fig. G.1 Energy band diagram of a p-type substrate at flat band. (After [2].)

$$f_{FD} = \frac{1}{1 + \exp[(E_t - E_F)/kT]}, \quad (\text{G.1})$$

where E_t is the energy level of the interface trap and E_F is the Fermi energy level (Fig. G.1). The probability of occupation of an energy state at the Fermi level energy is one-half.

Defining the potential of the interface trap with respect to the intrinsic energy level E_{is} at the surface as

$$E_t - E_{is} = -q\phi_t, \quad (\text{G.2})$$

the Fermi-Dirac distribution becomes

$$f_{FD} = \frac{1}{1 + \exp\{-(u_s + u_t) - u_F\}} \quad (\text{G.3})$$

where the variable u represents the potentials normalized to the thermal voltage. Out of equilibrium the occupancy of the interface traps is given by the Hall-Shockley-Read statistics (see Appendix F), according to expression (F.11), which is simplified to

$$f_{HSR} = \left[1 + \frac{p + n_i \exp(-u_t)}{n + n_i \exp(u_t)} \right]^{-1} \quad (\text{G.4})$$

for equal capture cross-sections of electrons and holes. For a p-type semiconductor, the electron and hole densities are given by (2.4.9) and (2.4.10) repeated below

$$p = n_i e^{u_F - u} \quad (\text{G.5})$$

$$n = n_i e^{u - u_F - u_C}. \quad (\text{G.6})$$

At the surface of the semiconductor, $u = u_s$; thus, the substitution of (G.5) and (G.6) into (G.4) yields

$$f_{HSR} = \left[1 + \frac{\exp(u_F - u_s) + \exp(-u_t)}{\exp(u_s - u_F - u_C) + \exp(u_t)} \right]^{-1}, \quad (\text{G.7})$$

which can be written as

$$f_{HSR} = \left\{ 1 + \left[\frac{\exp(u_F - u_s) + \exp(-u_t)}{\exp(u_F - u_s + u_C) + \exp(-u_t)} \right] \right. \\ \left. \times \left[\exp\{-(u_s + u_t) - (u_F + u_C)\} \right] \right\}^{-1}. \quad (G.8)$$

A detailed analysis of (G.8) is presented in [2], where it has been shown that for small channel voltages and high inversion level,

$$\frac{\exp(u_F - u_s) + \exp(-u_t)}{\exp(u_F - u_s + u_C) + \exp(-u_t)} \cong 1. \quad (G.9)$$

Therefore, (G.9) reduces to

$$f_{FD} = \left[1 + \exp\{-(u_s + u_t) - (u_F + u_C)\} \right]^{-1}, \quad (G.10)$$

where we have written f_{FD} instead of f_{HSR} . In effect, (G.10) is similar to (G.3) with $u_F + u_C$ in the place of u_F . Thus, (G.10) can be viewed as a quasi-Fermi-Dirac distribution function, in which the quasi-Fermi level of the electrons substitutes the Fermi level of equilibrium. It can be shown [2] that (G.10) gives a good approximation over the whole operating region of the transistor, provided that the drain voltage is limited to about one volt.

Considering the zero Kelvin approximation of the Fermi-Dirac distribution, that is,

$$f_{FD} = 1 \text{ for } E_t \leq E_{Fn} \quad (\phi_s + \phi_t \geq \phi_F + V_C) \\ f_{FD} = 0 \text{ for } E_t > E_{Fn} \quad (\phi_s + \phi_t < \phi_F + V_C) \quad (G.11)$$

(G.11) states that all the surface states below the quasi-Fermi level of minority carriers are occupied by an electron while the surface states above are unoccupied.

Now that the statistics of occupation of the surface states has been determined, let us proceed to calculate the effect of the surface states on the MOS capacitance. Let us first recall that the effective fixed oxide charge per unit area contains a component Q'_{ss} that represents the interface charge. The presence of interface states will be responsible for changes in the interface charge in response to the applied voltages. Let $\Delta Q'_{ss}$ be the supplementary interface charge with respect to the flat-band

condition. For a normalized band bending potential equal to u_s and a splitting of the quasi-Fermi potentials equal to u_c , the supplementary interface charge is written as

$$\Delta Q'_{ss} = \int_{E_V}^{E_C} N'_{ss}(E) f_{FD}(E) dE \Big|_{u_s} - \int_{E_V}^{E_C} N'_{ss}(E) f_{FD}(E) dE \Big|_{u_s=0}. \quad (G.12)$$

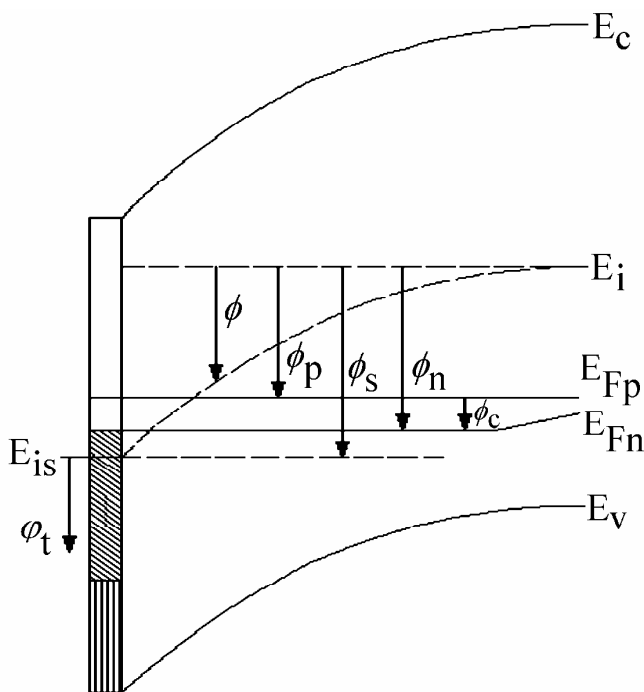


Fig. G.2 Energy band diagram for $\phi_s=440\text{mV}$ and $V_c=80\text{mV}$. (After [2].)

Using the Fermi-Dirac statistics in (G.11) and assuming that N'_{ss} is constant, (G.12) is simplified to

$$\begin{aligned}
 \Delta Q'_{ss} &= N'_{ss} \left[\int_{E_V}^{E_{Fn}} dE \Big|_{u_s, u_C} - \int_{E_V}^{E_F} dE \Big|_{u_s=u_C=0} \right] \\
 &= N'_{ss} \left[E_V(u_s=0) - E_V(u_s) + E_{Fn} - E_F \right] \quad (G.13) \\
 &= -qN'_{ss} [\phi_s - V_C].
 \end{aligned}$$

The meaning of the potentials and energies in expression (G.13) are explained in Fig. G.2. The interface trap capacitance C'_{ss} is then given by

$$C'_{ss} = -\frac{d\Delta Q'_{ss}}{d\phi_s} = qN'_{ss}. \quad (G.14)$$

Bibliography

- [1] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, New York, 1967.
- [2] Paul A. Muls, Gilbert J. Declerck, and Roger J. Van Overstraeten, "Characterization of the MOSFET operating in weak inversion", pp. 197-266 in *Advances in Electronics and Electron Physics*, vol. 47, Academic Press, San Diego, 1978.

This page intentionally left blank

Index

A

Acceptors, 391
Accumulation region, 13
ACM model, 370-373
Avalanche phenomenon, 309

B

Backscattering coefficient, 130
Ballistic transport, 129-131
Benchmark tests for dc MOSFET models
 Gummel symmetry test, 149-150
 series-parallel association of
transistors, 145-147
Body effect factor, 23
Boltzmann's constant, 8
Breakdown, 313
BSIM5 model, 375-376
Bulk charge density, 41

C

Capacitance
 bulk-to-drain, 177
 bulk-to-source, 177
 drain-to-source, 177
 gate-to-bulk, 179
 gate-to-drain, 176
 gate-to-source, 176
 extrinsic, 188
Capacitance per unit area
 depletion, 39
 flat-band, 21
 gate-bulk, 18
 inversion, 39
 oxide, 17
 interface trap, 52
Capacitor, metal-oxide-semiconductor (MOS), 14

Carrier, concentrations, 7
Channel-length modulation (CLM), 128-129
Channel segmentation, 280
Charge density
 electron, 39
 hole, 39
 volumetric, 391
Charge per unit area
 depletion, 39
 inversion, 39
 semiconductor, 17
Charge sharing, 132
Charge-sheet approximation, 41, 81, 85
CLM (channel-length modulation), 128-129
Compact models, 2
Conductance, output, 141
Contact potential, 26
Continuity equation, 158, 397-398
Corner frequency, 250
Correlation coefficient between drain and gate noise, 288-289
Coulomb scattering, 114
Critical field, 119
Cutoff frequency, 185

D

Debye length
 extrinsic, 392
 intrinsic, 339
Depletion capacitance, 39
Depletion region, 13
DIBL (drain-induced barrier lowering), 137-139
Dielectric constant, 387
Diffusion, 33, 394
Diffusion constant, definition, 394

Donors, 391
Doping, 391
Drain current, 73-78
Drain current, effect of velocity saturation, 122
Drain stored charge, 162, 167
Drain-induced barrier lowering (DIBL), 137-139
Drift, 33, 393
Drift velocity, 119

E

Early voltage, 141
Effective channel length, extraction, 355
Effective channel width, extraction, 355
Einstein relationship, 33, 395
EKV model, 373-375
Electric field in semiconductor, 12
Electron and hole concentrations, equilibrium, 7
non-equilibrium, 33
Electrostatics, 387
Energy bands in semiconductor, 11
Equivalent density of oxide traps, 239
Equivalent impurity density for mismatch calculation, 210
Extrapolated threshold voltage, 350-351
Extrinsic capacitances, 188-189

F

Fermi potential, 26
Fermi-Dirac statistics, 409-410
Field effect in bulk semiconductors, 9
Flat-band condition, 21
Flat-band voltage, 29-31
Flat-band voltage, effects of work function difference and oxide charges, 29
Flicker noise corner frequency, 250
Flicker noise in MOSFETs, 237-240
Flicker noise model in terms of inversion level, 247-248
Forward saturation current, 88-89
Fowler-Nordheim tunneling, 296

G

Gate current partition into source and drain components, 303-304
Gate transconductance, 98
Gate-induced drain leakage (GIDL), 306
Gauss' law, 387
GIDL (gate-induced drain leakage), 306
Gradual-channel approximation, 74

H

Hall-Shockley-Read statistics, 405-407
Halo-implanted MOSFET, 321, 323
High-frequency small-signal model, 277-278
HiSIM model, 378-379
Holes, 7

I

Impact frequency, 297
Impact ionization current, 308
Impedance field method
long-channel MOSFET, 233-234
short-channel MOSFET, 253-255
Induced gate noise, 281-289
Interface trap capacitance, 409-413
Interface traps, 237-238, 405
Intrinsic carrier concentration, 8
Intrinsic cut-off (transition) frequency, 185
Intrinsic small-signal models, 173
Intrinsic transition (cut-off) frequency, 185
Inversion capacitance, 39
Inversion charge, 167
Inversion charge density, 41
Inversion coefficient, 89
moderate inversion, 90
strong inversion, 90
weak inversion, 90
Inversion level, 89

J

Junction breakdown, 313
Junction capacitance, 188
Junction, pn, 399-403

K

Kingston function, 37

M

Mass-action law, 8

Mathiessen's rule for mobility, 114

Mismatch model

- equivalent impurity density, 210
- in terms of inversion level, 213-215
- number fluctuation, 207-211
- specific current mismatch, 212

Mobility, 33

Mobility, definition, 394

Mobility, effective, 117-119

Moderate inversion, 13, 90

MOS capacitor

- effects of interface traps, 52
- effects of measurement frequency

on the C-V curve, 25

- effects of polysilicon gate, 54-57

ideal, 14-17

ideal C-V plot, 21-22

quantum mechanical effects, 57-61

small-signal equivalent capacitive

circuit, 17-19

surface potential versus gate

voltage, 20

MOS model 11, 380-382

MOS structure, three-terminal, 32

MOSFET capacitive coefficients

definitions, 171

effects of velocity saturation, 195-197

effects of DIBL, 199

formulas, 181

MOSFET extrinsic capacitances, 188-189

MOSFET model, low-frequency small-signal, 103

MOSFET output conductance, 141

MOSFET, quasi-static small-signal model, 175, 180

Multiple-gate transistors, 343

N

Narrow-channel effects, 135-136

Noise

corner frequency, 250

design-oriented noise models, 246-249

flicker noise, 231-232

flicker noise in MOSFETs, 237-240

shot noise, 229-231

thermal noise, 227-229

thermal noise in MOSFETs, 236

Non quasi-static operation, 266

Non quasi-static small-signal model, 267-275

Non quasi-static y-parameter model, 277-278

Normalization (specific) current, 89-90

Normalization (specific) current, extraction, 352-354

Number fluctuation mismatch model, 207-211

O

Output conductance, 141

Oxide charges, 29-31

P

Pao-Sah model, 73-78

Parameter extraction, 349-350

Permittivity, 387

Phonon scattering, 114

Pinch-off voltage

bulk transistor, 45

SOI transistor, 334

pn junction, 399-403

Pocket-implanted MOSFET, 321-322

Poisson equation, 388

Poisson-Boltzmann equation in bulk semiconductors, 10

Polysilicon gate, 54

Power spectral density, 231

Q

Quantum-mechanical effects, 57-61

Quasi-Fermi potential, 33

Quasi-Fermi potentials in a pn junction, 402

Quasi-static approximation, 162

R

Recombination, 397

Relationship between inversion charge density and surface potential, 43-44
Resistances, series, 129
Reverse saturation current, 88-89
Reverse short-channel effect, 134

S

Saturation charge, 123
Saturation coefficient, 165
Saturation current
 forward, 89
 reverse, 89
Saturation level, 96
Saturation region, 89, 92-93, 96
Saturation velocity, 119
Saturation voltage, 96, 126
Semiconductor capacitance, 18
Short-channel effects, 131-133
Short-channel parameter, 122
Shot noise, 230
Silicon-on-insulator (SOI) MOS transistor, 325
Slope factor, 43
Slope factor, extraction, 357-358
Small-geometry effects, 140-144
Small-signal model, 103-105
Small-signal transconductances, 98
SOI transistor model, 327-334
Source stored charge, 161, 167
SP model, 382-383
Specific (normalization) current, 89-90
Specific (normalization) current, extraction, 352-354
Standard deviation, 210
Strong inversion region, 13, 90
Substrate (bulk) current, 306
Substrate current-induced body effect (SCBE), 315
Subthreshold (weak inversion) region, 91-93
Subthreshold slope, extraction, 358-359
Surface electric field, 37
Surface potential, 16
Surface potential master equation, 37
Surface roughness scattering, 114

T

Temperature effects, 116, 139

Thermal charge, 45
Thermal noise
 resistor, 228-229
 long-channel MOSFET, 236
 short-channel MOSFET, 255-258
 gamma factor, 259

Thermal voltage, 9

Three-terminal MOS structure, small-signal capacitive model, 40

Threshold voltage, 45, 46
 extraction, 352-354
 extrapolated, 350-351
 narrow-channel effect, 136
 reverse short-channel effect, 134
 short-channel effect, 132
 temperature dependence, 139-140

Transconductance

 source, 98
 gate, 98
 bulk, 98
 drain, 98

Transconductances, 98

Transconductance-to-current ratio, 101

Transit time, 170

Transition frequency, 185

Tunneling current, 299

Tunneling through a potential barrier, 294

 Fowler-Nordheim tunneling, 296
 direct tunneling, 296

U

Undoped surrounding-gate transistor model, 338-343

Unified charge control model (UCCM), 47-48, 90

V

Velocity saturation, 119-122

Virtual charge density, 190

Volumetric charge density, 391

W

Weak avalanche, 309-311

Weak inversion region, 13, 90

Wentzel-Kramers-Brillouin (WKB) approximation, 295

Work function, 26



Carlos Galup-Montoro studied engineering sciences at the University of the Republic, Montevideo, Uruguay, and electronic engineering at the National Polytechnic School of Grenoble (INPG), France. He received an engineering degree in electronics in 1979 and a doctorate degree, for a study on short channel MOS transistor degradation, in 1982, both from INPG.

From 1982 to 1989, he was with the University of São Paulo, Brazil, where he was engaged in junction field effect transistor (JFET) fabrication and analog circuit design. Since 1990, he has been with the Electrical Engineering Department, Federal University of Santa Catarina, Florianópolis, Brazil where he is now a professor. There he has been involved in the design of analog circuits fully compatible with digital technology and low voltage analog circuits, and in device modeling. From August 1997 to February 1998 he was a research associate with the Analog Mixed Signal Group, Texas A&M University.

His main research interests and expertise are in field effect transistor modeling, and transistor-level design.



Márcio Cherem Schneider received the BEng and MSc degrees in Electrical Engineering from the Federal University of Santa Catarina (UFSC), Florianópolis, SC, Brazil, in 1975 and 1980. In 1984 he received the PhD degree in electrical engineering from the Universidade de São Paulo, Brazil, for a dissertation on bipolar transistor modeling. In 1976, he joined the Electrical Engineering Department at UFSC, where he is now a Professor. In 1995, he spent a one-year sabbatical at the Swiss Federal Institute of Technology (EPFL). In 1997 and 2001, he worked as a Visiting Associate Professor with the Department of Electrical and Computer Engineering at Texas A&M University. He has ongoing research programs in the areas of analog integrated circuits, low-voltage/low-power circuits, and MOS transistor compact modeling which have resulted in more than 60 publications in international conferences and journals. His current research interests are mainly focused on transistor modeling and transistor-level design, in particular of analog and RF circuits.